Enhancing Accessibility to Heterogeneous Sri Lankan Cultural Heritage Information across Museums through Metadata Aggregation

Enhancing Accessibility to Heterogeneous Sri Lankan Cultural Heritage Information across Museums through Metadata Aggregation

Student No: 201425006
Name: Chiranthi Jayampathini Wijesundara

Cultural Heritage Information (CHI) is an essential resource which exhibits values of a society. Memory institutions play the main role of delivering CHI to the public. This study focuses on CHI of Sri Lankan cultural heritage collected by museums. Museums usually handle heterogeneous information compared with other memory institutions. Due to these heterogeneity museums tend to adopt unique standards according to their institutional requirements. Developing countries like Sri Lanka still does not possess strong CHI delivery portals for cultural objects and artefacts within the country, and their standards are still under development. Nevertheless, museums outside Sri Lanka that own Sri Lankan cultural objects provide valued CHI that can be retrieved through online collections. This study sought to find an approach to aggregate Sri Lankan CHI across museums in and out of the country and deliver them to the patrons with more contextual information. The study also seeks a method to eliminate the disparity in museum standards through a metadata crosswalk approach between museum vocabularies. This target was achieved by investigating over 2600 object records across four museums, namely, British Museum, Victoria and Albert Museum, Metropolitan Museum of Art and a teaching museum attached to the University of Peradeniya, Sri Lanka. The mapping was based on the *object categories* of the museum objects and the key vocabulary used was the Getty Art and Architecture Thesaurus (AAT). In this thesis, the mappings are presented as RDF graphs to show the relationships between the AAT terms and the museum vocabularies. The metadata-level aggregation models were developed to show the relationships through *spatial*, *temporal* and *thematic* terms related to the cultural objects and the information was enriched through Linked Open Data (LOD) resources. The final outcome of the research was a metadata model which aggregates Sri Lankan CHI. The main platform of this aggregation model depended on the vocabulary crosswalk approach mentioned above. The resulting mapping derived trough the crosswalk provided enhanced meaning to the cultural objects and the same approach can be extended to develop more comprehensive level metadata vocabulary mapping and metadata aggregation across Sri Lanka and South-East Asian memory institutions in the future.

Academic Advisors: Principal: Shigeo SUGIMOTO
Secondary: Mitsuharu NAGAMORI

Enhancing Accessibility to Heterogeneous Sri Lankan Cultural Heritage Information across Museums through Metadata Aggregation

Chiranthi Jayampathini Wijesundara

Graduate School of Library, Information and Media Studies
University of Tsukuba

August 2016

Table of Contents

List of Figures

List of Tables

Appendices

1. Introduction

Cultural heritage is a showcase of any society which represents unique characteristics and values related to a community. These cultural heritages can be found all over the world and they are preserved and exhibited or kept in museums, libraries or archives. Some of these heritage collections can be retrieved via the Internet. In this research, the main area of study is Cultural Heritage Information (CHI) of museums with special reference to Sri Lanka. During the colonial period, foreigners took Sri Lankan artefacts to their countries and they were in their private collections. Later many of these artefacts were donated or brought back by the museums. At present these objects can be viewed online through relevant institutional data portals and they use different metadata standards to organise these CHI collections.

This diversity of metadata standards which use to organise the CHI by these memory institutions, creates metadata interoperability issues leading to poor networking amidst museums. On the other hand, the museum collections on the internet are operated as isolated portals and users have to access each portal individually to find a certain information. The information provided by these online collections are mostly context neutral and frequently information users need more descriptive and contextual information to fulfil their information needs. Besides, Sri Lanka does not possess an appropriate way of connecting CHI scattered all over the internet which is very useful when linking, and searching heritage information within a common container.

Pointing out above research problems, the study aims to design a platform to connect and aggregate CHI related to Sri Lanka in different memory institutions through a semantic metadata model. The proposed model collects metadata from different individual museums and aggregates this heterogeneous CHI while enriching the contents of the same. Secondly, the study tries to identify the metadata related to Sri Lankan cultural heritage objects in museum collections. To fulfil this objective, museum information from Europe, North America and some offline Sri Lankan museums were investigated and selected terms were mapped to Getty AAT vocabulary to make a formalisation. Besides, this study aims to incorporate appropriate data standards to standardise the above model while integrating LOD (Linked Open Data) technologies and enrich the information contents and deliver more contextual information to the users.

Somehow, there are efforts on designing data portals to integrate digital cultural heritage into a single platform and enrich these cultural heritage objects more meaningful through Linked Open Data technologies. One such example is the Project Europeana designed for European Union Countries. The main aim of the project Europeana is to collect metadata

from a large number of providers, mainly cultural institutions, across Europe, and to enable search and discovery of cultural items described therein (Haslhofer & Isaac, 2011). Nevertheless, identifying different metadata in diverse CHI collections and developing an aggregation model to connect solely Sri Lankan CHI would be a novel attempt for Sri Lanka domain. Besides, developing an own model or custom-made model would be an ideal solution to eliminate specific context integration issues during data aggregation. Usually, most tailor-made models cannot fit into a diverse and unique data aggregation and such aggregation might result in information loss or omission of information. Hence, this study provides a solution to overcome such issues.

The methodology of this study can be summarised as follows. First, the data is collected through three museums covering Europe and the North America. Then the attributes are carefully investigated and selected terms were mapped to Getty AAT vocabulary to form a unified mapping between *object categories* of the heritage objects. Some offline museum data related to a local Sri Lankan museum was collected and mapped similarly. Later, few CHI instances were selected and developed the model through RDF (Resource Description Framework) technology. However unlike well designed renowned data models, the approach used in the study can be identified as a bottom-up approach because it starts from the base level metadata aggregation. Since there was no solid data provider the metadata were collected through online collections scattered on the Internet. Somehow, the relationships were established based on specific attributes of the developed generalised aggregation model which was derived from the diverse museum collection schemas.

The beneficiaries of this research are mainly the cultural heritage information users and related organisations. Similarly, the proposed aggregation model can be used by Web portal developers as a base model to develop a cultural heritage portal for Sri Lanka or South East Asia.

The organising of the thesis is mainly based on six chapters and it can be described in the following manner. Chapter one is dedicated to the introduction of the thesis and it includes an outline of the research as a whole. The study area, research problems, objectives, significance of the study, and methodology are described briefly in this section. Second and third chapters are dedicated to defining the specific terms and related literature and theoretical aspects of the study. Chapter four represents the research methodology and related techniques. This section answers the question of how the target study is carried out. Chapter five describes the results and discussion of the study. Final or the sixth chapter is dedicated to discussing the conclusions and future directions of the study.

2. Cultural Heritage Information (CHI) and Museum Data

2.1. Cultural Heritage Definition

With regards to the Cultural Heritage first we should identify what is meant by 'Heritage'. "Heritage is a broad concept and includes the natural as well as the cultural environment. It encompasses landscapes, historic places, sites and built environments, as well as biodiversity, collections, past and continuing cultural practices, knowledge and living experiences" (ICOMOS, 2002). According to the UNESCO heritage can be categorised as follows.

i. Cultural heritage
    a) Tangible cultural heritage:
    - Movable cultural heritage (paintings, sculptures, coins, manuscripts)
    - Immovable cultural heritage (monuments, archaeological sites etc.)
    - Underwater cultural heritage (shipwrecks, underwater ruins and cities)
    b) Intangible cultural heritage: oral traditions, performing arts, rituals
ii. Natural heritage: natural sites with cultural aspects such as cultural landscapes, physical, biological or geological formations
iii. Heritage in the event of armed conflict (UNESCO, 2016)

According to the above classification, it is clear that Tangible and Intangible cultural heritage goes under the category 'Cultural Heritage'. UNESCO further describes the Cultural Heritage as, "…. the legacy of physical artefacts and intangible attributes of a group or society that are inherited from past generations, maintained in the present and bestowed for the benefit of future generations" (UNESCO, 2016).

Under the division of Cultural Heritage, the 'Tangible' cultural heritage is further subdivided into three categories. Since here the investigation is about museum data, here in this research the main focus is on the 'Movable Cultural Heritage' which comes under the 'Tangible Cultural Heritage' category. That means all the tangible and portable objects which we can find in museums or archives such as, paintings, sculptures, vessels, manuscripts are come under the above category

Somehow it is essential to understand the value and the meaning of this cultural heritage in parallel with the society they belong to. In that sense museum plays a major role in collecting, preserving, interpreting, and displaying items of artistic, cultural, or scientific significance for the education of the public.

2.2. CHI and Metadata Standards

As identified earlier Cultural heritages can have various categories. Scholars discuss a new form of cultural heritage when the tangible or intangible heritage objects are digitised; and this is commonly referred to as digital cultural heritage or cultural heritage information resources (Lor and Britz, 2012). However, cultural heritage possesses a vast number of information resources. They can be mainly divided into digital, non- digital and born- digital resources. According to the Figure 1 these CHI resources can be varied from a printed book to a born digital virtual reality programme.

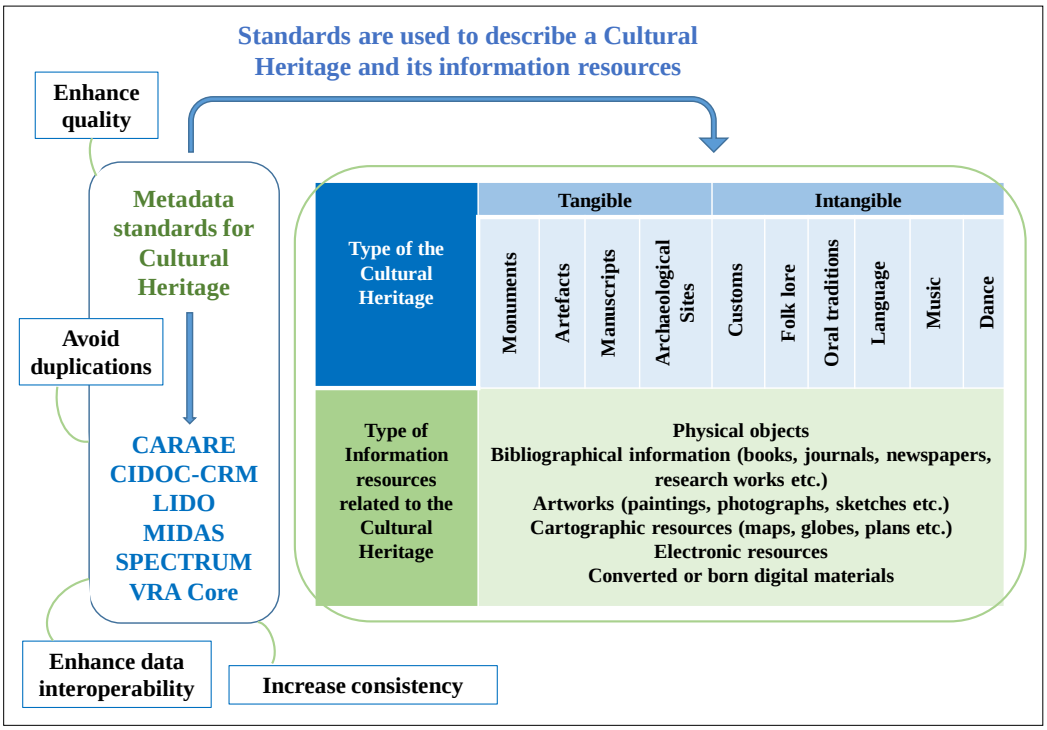


Figure 1: Different types of cultural heritage objects, their information resources and its connection between the metadata standards

When dealing with any information, having an accepted standard is essential because it affects the longevity, quality and interoperability of the information. Therefore, the above requirements can be fulfilled by involving the metadata standards into the CHI process (Figure 1). Metadata, literally known as “data about data” is widely used CHI professionals to create value-added information and such metadata is often governed by well-known standards and best practices in order to ensure the quality, consistency, and the interoperability of data (Gilliland, 2008, p. 1). According to the Figure 2 CHI lifecycle can have five major phases.

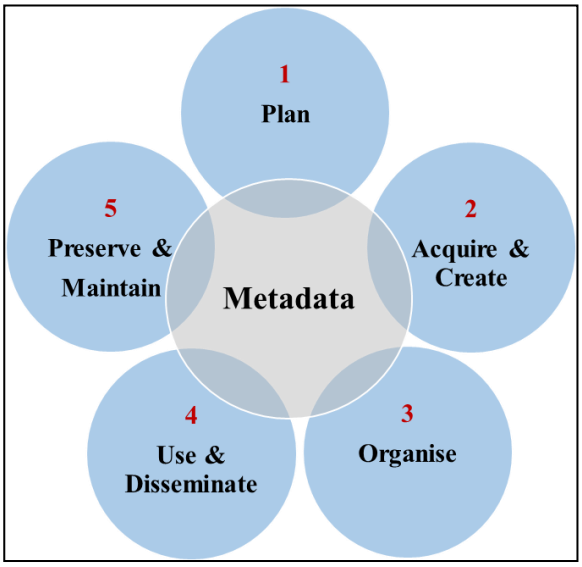


Figure 2: Lifecycle and involvement of metadata in CHI process

First heritage institutions should prepare a long-term plan explaining all the requirements, resources, techniques, data, metadata, risks and benefits of the process and how it should be carried out etc. Then in the second step, the institution should obtain data through other institutions and also they can create their own data. In this phase metadata involvement is essential because institutions must follow accepted cultural heritage standards during recording, creating and capturing right CHI data. Organising CHI data is the third phase and here also CHI metadata standards should be incorporated. Then the organised data should be utilised and disseminate through data portals. Finally, the maintenance and preservation of CHI are critical to ensure the long-term use and existence of the data. Somehow during every phase of the CHI processes different metadata types such as descriptive, administrative or structural metadata and related metadata standards can be used, created and utilised. The book *Introduction to Metadata* by J. Paul Getty Trust explains the need for involvement of metadata into memory institutions as follows.

> "..… institutions need to change old paradigms and procedures. They need to make a lasting commitment to creating and continually updating the various types of core metadata relating to their collections and the digital surrogates of collection materials that we all seem to be in such a hurry to create" (Baca, 2008, p. v).

2.2. Cultural Heritage in Sri Lanka

Sri Lanka which was known as Ceylon during British occupation is an island situated in the Indian Ocean. Despite the size of the country, eight world-class heritage sites are situated in Sri Lank. Apart from these there are many local heritage sites scattered all over the country. These assets are one of the major income-generating resources in the country through the generation of tourism.

The chief regularity body who is responsible for the cultural heritage sites and monuments of Sri Lanka is the Department of Archaeology. The museums are governed by the Department of National Museum. The Central Cultural Fund, the Department of Cultural Affairs, National Archives and National library are also responsible for the activities related to the cultural heritage of the country.

2.2.1. Metadata Approaches in Sri Lankan Cultural Heritage

The National Library of Sri Lanka and University libraries' efforts on digitising bibliographical materials can be shown as initial attempts of using metadata standards into Sri Lankan heritage sector. Even though their main concern is on bibliographical materials they use standards such as DDC, LCSH, MARC, Dublin Core etc. to organise their information. For instance, National Library is committed to maintaining appropriate standards for managing information materials.

> "To maintain long-term preservation that means the accessibility to the digitized items will depend on suitable standards. Accordingly, the library will adhere to established internationally accepted standards" (Gangabadadarachchi & Amarasiri, 2009).

Palm Leaf Manuscript Study and Research Library is a digital library project carried out by the University of Kelaniya, Sri Lanka. Under this project, the library preserves palm leaf manuscripts found in Sri Lanka through digitization and provides access to them in order to facilitate study and research in the sphere of manuscript culture (Ranasinghe, 2015).

In addition, Sri Lanka National Library and Documentation Services Board (NLDSB) joined with IFLA (International Federation of Library Associations) and formed an institute called *IFLA Preservation and Conservation (PAC) Centre* in the year 2015 (Cabral, 2016). IFLA-PAC Centre aims to preserve ancient writings such as Palm Leaf Manuscripts and they coordinate UNESCO Memory of the World (MOW) Program and UNESCO Intangible Cultural Heritage (ICH) Program which are highly related to Sri Lankan CHI sector.

Although Sri Lankan libraries and archives consider and practice metadata standards, the cultural heritage institutions such as museums poorly incorporate them into their fields.

Often they rely on their own standards and they record and organise the CHI data according to the organisations' standards. Also a handful of research have been addressing these metadata issues and one such example is the paper presented by the author at the DOCAM Conference at Sydney in July 2015. The title of the paper was *Documenting Spatial and Temporal Information for Heritage Preservation: A Case Study of Sri Lanka.* This paper proposed that with recent developments in the field of Geographic Information Systems (GIS) and Remote Sensing (RS), heritage preservation can be enhanced and improved by documenting Spatial and Temporal (ST) information parallel to the other information. The paper investigates the present condition of the ST information in the heritage arena and the challenges associated with the same. The study area was heritage sites in Sri Lanka and at the end the paper, authors proposed a basic metadata model, which can be used to acquire Spatial and Temporal information during archaeological site recording. This study utilised renowned cultural heritage standards such as MIDAS Heritage standard of UK to design the model and this can be shown as a good practice of data acquisition in the CHI documentation arena (Wijesundara, Sugimoto, & Narayan, 2015).

2.3. Museum Information

According to ICOM "a museum is a non-profit, permanent institution in the service of society and its development, open to the public, which acquires, conserves, researches, communicates and exhibits the tangible and intangible heritage of humanity and its environment for the purposes of education, study and enjoyment" (ICOM, 2007).

Similar to a library or archives, the museum also holds a massive amount of information related to cultural heritage. However, there is a major contrast between a library and a museum resources. Museums usually hold heterogeneous objects or artefacts. Often these artefacts are unique and exist as sole objects. The museums organise these artefacts using different standards and meanwhile deliver them to the patrons through online collections. Following figure (Figure 3) shows an example of a museum object and its associated CHI in their collection on the web. The example shows essential CHI which is used to describe a *comb* and the highlighted red coloured words give links to further explanations (described as *Scope Note*) designed by the British Museum, UK.

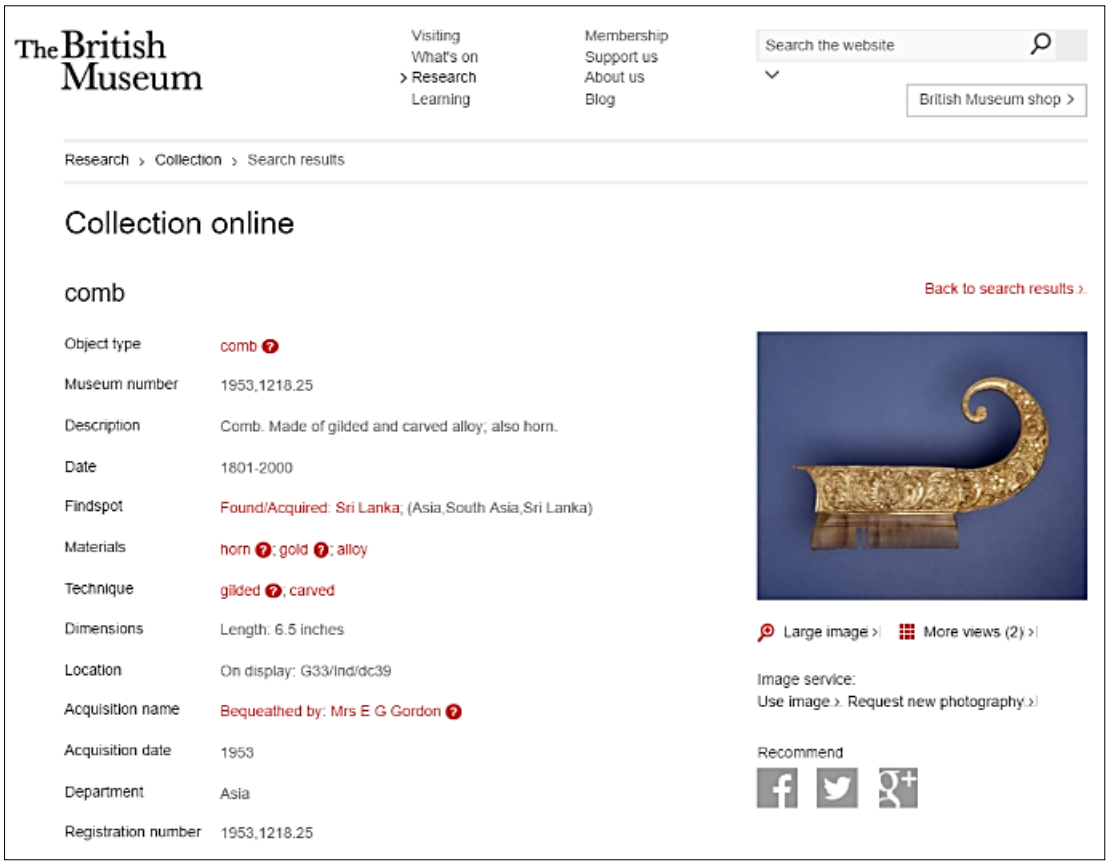


Figure 3: A museum object with its related CHI retrieved from the British Museum collection online (British Museum, 2016)

2.3.1. Metadata Standards Associated with Museum Data

Prior to creating online collections, museums had conventional card catalogue systems to organise their unique information. They were similar to traditional library catalogues and museums used accepted museum standards to categorise their CHI. Museum standards are an essential reference for museums of all kinds.

Museum standards guide towards:

- effective management
- appropriate care of collections
- connecting people with collections and cultural heritage (Museums Australia- Victoria, 2016).

The main body who deals with the museums is the ICOM or the International Council of Museums. The initiating of ICOM goes back to 1946 and it sets standards for museums in

design, management, collection organisation and maintains ICOM Code of Ethics for museums etc. (ICOM, 2016).

There are many museum standards world over and the following table (Table 1) shows some of those standards and their corresponding organisations and related information as follows.

Table 1: Diverse museum standards around the world

	Standard	Description	URL
01	AFRICOM Standard	AFRICOM is an organisation to promote museums standards and code of ethics in Africa. AFRICOM Handbook of Standards developed by ICOM and the AFRICOM Co-ordinating Committee for use by museums throughout Africa in 1996.	http://archives.icom.museum/africdoc/light/index.html
02	CIDOC Conceptual Reference Model (CRM)	This was initiated in the early 1990s by the ICOM/CIDOC Documentation Standards Group. The latest version of CIDOC-CRM is version 6.2.1 which was realised in year 2015. This standard provides an extensible ontology for concepts and information in cultural heritage and museum documentation.	http://www.cidoc-crm.org/
03	Dublin Core (DC)	Started in 1995 in Dublin, Ohio. This consists of 15 core elements and was extend to a larger vocabulary through Dublin Core Metadata Initiative (DCMI). This standard is usable for describing a wide range of resources in diverse communities.	http://dublincore.org/documents/dces/
04	MDA (Museum Documentation Association) Data Standard	MDA Data Standard was initiated in year 1991. MDA was re-launched as the Collections Trust in 2008 at London.	http://www.collectionstrust.org.uk/about-collections-trust/history

05	Object ID	The international standard, Object Identification (Object ID), developed by the Getty Information Institute in 1997. This gives essential information about archaeological, artistic and cultural objects in order to facilitate their identification in case of theft.	http://archives.icom.museum/objectid/
07	SPECTRUM	SPECTRUM is the UK Museum Collections Management Standard. The first edition of SPECTRUM was developed in 1994	http://www.collectionstrust.org.uk/spectrum
07	VRA Core (Virtual Resources Association)	VRA Core is a data standard for the description of works of visual culture as well as the images that document them. This is s hosted by the Library of congress (LC).	http://www.loc.gov/standards/vracore/

Apart from the above main standards there are local standards that are specific to countries such as, National Standards for Australian Museums and Galleries, American Alliance of Museums (AAM) Standards, Structured Model for Museum Object Information by Tokyo National Museum, National Standard for Cultural Property of South Korea, Cataloguing Cultural Objects (CCO) by Princeton University Art Museum etc.

Controlled vocabularies and thesauruses are another special kind of standard to represent terms and concepts of a specific field. They are an essential component when it comes to enriching and aggregating metadata in different institutions. According to the Getty Institute "…. purpose of controlled vocabularies is to organize information and to provide terminology to catalog and retrieve information. While capturing the richness of variant terms, controlled vocabularies also promote consistency in preferred terms and the assignment of the same terms to similar content" (Harpring, 2010, p.12). There are specifically designed thesauruses and vocabularies for museum environment also.

E.g. Getty AAT (Art and Architecture Thesaurus)/ Getty CONA (Cultural Objects Name Authority)/ Getty TGN (Thesaurus of Geographic Names)/ British Museum Object Names Thesaurus/ British Museum Materials Theseus etc.

Therefore, it is obvious that there are plenty of museum standards used by memory institutions around the world. The main reason for these diverse standards is the heterogeneity and the uniqueness of museum CHI which causes difficulties to organising them into a single

platform. This issue and possible solutions for it would be further discussed in the third section of this thesis.

According to Murtha Baca et al. “No single schema or controlled vocabulary is likely to answer all the needs of any institution……every institution will be required to piece together its own metadata and cataloguing strategy from the available options……However, it is becoming clear that carefully crafted, standards-based……. metadata are a crucial part of any strategy aimed at creating interoperable, coherent, intelligible, and long-lived information sets” (Baca et al., 2008, p.126)

2.3.2. Sri Lankan Cultural Heritage in Museums Abroad

Before discussing the local museum information, let us look at the Sri Lankan artefacts in foreign museums.

Since Sri Lanka is an island and it is situated in an important place in the Indian Ocean it was often influenced by other countries. Throughout the historic period, some parts of the country were attacked and ruled by Tamil and South Indian rulers for short periods and during that time few invaders destroyed the historic monuments and however, most of them were rebuilt by the Sinhalese. Then after the 16th century, the country had to face the European invaders such as Portuguese, Dutch and British. Although Portuguese and Dutch ruling were confined to coastal areas of the country, in the year 1815 the British emperor managed to fully capture the whole country until the country regain independence in 1948 (De Silva, 1959).

During this colonial and Dutch ruling periods foreigners took many historical artefacts for their personal pleasure and as gifts to their friends. Conversely, during the British occupancy country’s history was reborn trough their efforts of discovering overlooked archaeological sites in Sri Lanka. British officials such as Mr H.C.P. Bell made some tremendous efforts to uplift the heritage of Sri Lanka. Though, there were no rules or regulations for exporting cultural artefact during that time colonial rulers took away an enormous amount of Sri Lankan artefacts to their countries.

After gaining the independence there were discussions and appeals to return those cultural objects back to Sri Lanka. During April 1980, UNESCO formed a committee called, *Intergovernmental Committee for Promoting the Return of Cultural Property to its Countries of Origin or its Restitution in case of Illicit Appropriation.* (UNESCO, 1980). Under that Sri Lanka also made a statement and according to that 27 countries and 140 institutions (mostly European institutions) owned artefacts belongs to Sri Lankan community. This statement was mainly

based on a book by Dr P. H. D. H. De Silva, and some of the prominent cultural objects in abroad can be shown as follows (Table 2).

Table 2: Cultural heritage objects in abroad (De Silva, 1975) & (UNESCO, 1980)

Country	Institution	Cultural Object
Austria	Kunsthistorisches Museum (Vienna)	- Ivory casket (16th century) - Ivory fan
Belgium	Musees Royaux d'Art et d' Histoire (Brussels)	- Knife - Ivory handle
Germany	Schatzkammer - Residenz (München)	- Two chest decorated with ivory, gold, rubies and sapphires (1545 A.D.) - Three carved combs with gold and rubies (1540 A.D.)
	Staatliche Museen, Museum Für Völkerkunde (Berlin)	- Around fifteen masks
Great Britain	Ashmolean Museum (Oxford)	- Carver ivory comb (18th century)
	British Museum (London)	- Bronze figure of god *Padmapani* - Bronze seated figure of *Cunda* (9th 10th century) - Bronze figure of seated *Avalokitesvara* (8th-10th century) - Ivory casket (1600 A.D.) - Silver ladle - Bronze gilt, standing figure of *Pattini* or *Tara* (10th century) - Sinhala sword (16th-17th century)
	British Museum Library	- Hugh Nevils' manuscripts collection - Two letters from king *Raja Simha* II (1652 & 1658 A. D.)
	Pitt Rivers Museum (Oxford)	- Carved ivory double-headed comb - A Flintlock gun (1740 A.D.)

	Royal Scottish Museum (Edinburgh)	- A bronze figure of seated Buddha - Bronze Buddha figure. seated on a coiled cobra (14th century)
	The Armouries H.M. Tower of London	- Elephant goad of exquisite craftsmanship with wooden handle (18th century)
	Victoria and Albert Museum (London)	- Cabinet of carved ivory with silver mounts 1700A.D.) - Carved ivory casket (17th-18th A.D.) - Bronze figure of Hanuman (11th century) - Circular brass dish (19th century) - Wooden casket - Ivory cabinet (18th century) - Silver waist chain (19th century) - A standing figure of Buddha in *Amaravati* style
France	Bibliothèque Nationale	- Palm leaf royal letters (1769 A.D.)
	Musée de I'Homme (Paris)	- Royal letter (1746 A.D.)
	Musée Guimet, place d'Iena (Paris)	- Two Ceylon masks
Netherlands	Rijksmuseum (Amsterdam)	- A gun (15th-16th A.D.) - Swords (16th century) - Daggers - Paintings (16th century)
	Rijksmuseum Voor Volkenkunde	- A bell with figures of *Hanuman* and *Garuda* - An exquisitely carved ivory pestle and mortar - Large ivory carving of a parrot
	Tropenmuseum (Amsterdam)	- Sinhalese Ola script (1785 A.D.) - Royal letter

	Tropenmuseum (Linnaeursstraat)	- Dextrally coiled, decorated conch shell
Switzerland		- *Kandyan* dagger - Lance-heads - Swords - Bronze figure of Buddha (18^{th}-19^{th} century) - Bronze statue of divinity
The United States of America	Boston Museum of Fine Arts	- A bronze figure of *Avalokitesvara* (8^{th} century) - A copper statuette of *Vajrapini* (9^{th} Century)
	Cleveland Art Museum, (Boulevard)	- Small bronze statue of Buddha (6^{th}-7^{th} century) - Bronze statue of deity (12^{th} century) - Carved ivory object

Currently, most of these world-class museum collections are online and as a result, people can witness those objects freely through the Internet. Apart from the above-mentioned institutions, there are many other memory institutions around the world which possess Sri Lankan heritage objects. Even though they belong to the Sri Lankan culture, public should not be worried because they are well preserved and well maintained by those prestigious memory institutions. Since they are well organised, presently Sri Lanka as well as other countries have convenient access to those museum collections for their study and research purposes.

2.3.3. Status of the Local Museums in Sri Lanka

When looking at the situation in Sri Lanka as mentioned once, the main institution responsible for movable artefacts is the Department of National Museums, Sri Lanka. Under the Department of National Museums, there are 10 main regional and special museums (Department of National Museums, 2016). Archaeological site museums are handled by the Department of Archaeology and the Central Cultural Fund, Sri Lanka. There are more than 25 of those and apart from that can find few private museums handled by various institutions as well. (Department of Archaeology, 2016). Basically, Department of National Museums

cooperates with the museum network in the island and it is closely coupled with the ICOM. Therefore, we can assume that the Sri Lankan museums are more or less standardised according to the ICOM standards. However, standards used by the Sri Lankan museums are not investigated under this study.

One of the major drawbacks of Sri Lankan museums is the absence of remotely access CHI. Unlike foreign museums, Sri Lanka does not provide online collections for the patrons. National museum and regional museums do maintain a standard manual recording system and a computer application to record information. Yet this information is strictly prohibited to the outsiders. During this research, the author too faced the same problem and finally had to depend on published printed catalogues. On the other hand, national museums do have published catalogues which are not restricted but, they do not cover all the objects in museums of Sri Lanka or at least the objects in the national museum, Colombo. Nevertheless, the main reason behind not releasing the CHI to the general public is due to unpredictable thefts in museums. In the year 2012 similar incident occurred and due to that museum lost some of the irretrievable treasures forever (Farisz & Dias, 2012). As a result, authorities are thrust to prevent developing an online museum portal, assuming that it will be an advantage for the museum robbers.

2.4. Goals of the Research in the Context of Sri Lankan CHI

As identified previously, it is obvious that Sri Lankan cultural objects are scattered all over the world and most of them can be retrieved individually through institutional portals. On the other hand, Sri Lanka still does not possess a CHI portal to deliver their information to the users. This can be identified as the main research problem in this study. The primary reason for this issue can be identified as the regional diversity. Compared to Sri Lanka, a top class museum in USA or Europe CHI possess well-managed authority controls, vocabularies and CHI are freely available as LOD. Yet due to various uncontrolled barriers such as data security, lack of technology and resources etc., Sri Lanka does not possess such strong vocabularies or LOD resources. In addition, the heterogeneity of cultural objects and diversity and needs of the memory institutions lead towards different metadata standards. Tony Gill (2004) discusses this diversity and its drawbacks as follows. "…these differences in descriptive schema across museums, libraries and archives, although necessary for individual applications, can seriously hinder cross–domain discovery and interoperability of cultural information resources in the global context of the Internet" (Gill, 2004).

Aiming those research problems, this study tries to find a solution to connect or aggregate scattered CHI on the web while trying to link the offline Sri Lankan information as

well. This aggregation is done through a metadata model specifically design for Sri Lankan CHI and this model can be used as a base model for future portal design or further semantic aggregations. Therefore, the main objective of this research is to aggregate diverse Sri Lankan CHI across memory institutions through a metadata model aiming a better accessibility to the information. Furthermore, the research tries to identify and map metadata of the CHI and tries to make a formalisation for the *object types* given by different museum collections. This formalisation is essential for semantic aggregation because the correctness of data integration is solely depending on that. Mapping the controlled vocabularies of museums to top class vocabularies such as Getty AAT can be identified as a unified approach for representing meta-metadata level CHI. As discussed earlier Sri Lanka does not possess its own CHI related vocabulary. Therefore, this kind of approach will be beneficial for a developing region such as Sri Lanka. Finally, the proposed aggregation model will be enriched using LOD resources aiming to facilitate data enrichment and to make CHI more contextual. It is a reality that even the top class museum collections do not provide many related details of their cultural heritage objects. Sometimes museums tend to omit certain information due to the lack of knowledge of individual communities and their unique vocabularies. To overcome this problem LOD enrichment will be an ideal solution.

However, metadata aggregation in CHI is not a novel approach to the world, but applying the same to a new domain like Sri Lanka will be a new avenue, which can be found in this study.

3. Literature on Metadata Aggregation and Related Approaches

3.1. Underlying Technologies and Concepts

This study mainly focusses on diverse metadata aggregation across museums. Therefore, it is essential to understand the related techniques and approaches which can be used in this study.

In this framework identifying methods to eliminate interoperability issues caused during the integration of diverse metadata standards, selecting the appropriate technology to describe the metadata models, and how metadata enrichment can be achieved should be investigated.

3.1.1. Metadata Crosswalk

To provide the user unified access to CHI need to develop a sophisticated tool which enables them to discover, access and share information across the collections. However, this can be only achieved through a formal metadata standard. As identified in the previous chapter diversity of standards used by the heritage institutions prevent creating such interface and it limits seamless access to CHI information. Similarly, it is clear that developing a formal metadata standard for museums is not a reality due to the diversity of the CHI they handle.

The ultimate solution to overcome this problem is metadata crosswalks. Term *Crosswalks* refers to the "…. mapping of the elements, semantics, and syntax from one metadata scheme to those of another" (NISO, 2004, p. 11)

The following figure (Figure 3) represents few instance of Dublin Core (DC) Metadata Element Set which was mapped into CIDOC Conceptual Reference Model (CRM) entities. According to the authors, metadata can be mapped into an ontology to provide interoperability of its data and to achieve information integration. Here DC can be identified as a metadata standard which is widely used while CIDOC CRM can be identified as an ontology in cultural heritage domain (Carrasco & Vidotti, 2015).

According to Woody (2008) by comparing two or more metadata schemas, similarities and differences can be identified and that enables to decide which schema is more interoperable and so on.

Table 3: Example showing DC to CIDOC CRM mapping (Carrasco & Vidotti, 2015).

Dublin Core	CIDOC CRM		Dublin Core	CIDOC CRM
Contributor	E39 Actor E74 Group E41 Appellation E10 Transfer of Custody E66 Formation		Type	E55 Type E17 Type Assignment
Coverage	E50 Date E52 Time-Span E53 Place E47 Spatial Coordinates E45 Address E48 Place Name		Publisher	E12 Production E29 Design or Procedure E51 Contact Point
Creator	E39 Actor E40 Legal Body		Identifier	E42 Object Identifier E15 Identifier Assignment

Another level of interoperability that can be achieved by crosswalks is the *Data Content Standards*. Here the *data values* are mapped and from that, a formal controlled vocabulary is created. Through this type of *integrated authority control* both interoperability and retrieval can be improved (Vellucci, 2011). In addition, if the data are in a single database crosswalk enables to search the whole collection, simultaneously by a single query (Zeng & Qin, 2015).

Conversely, crosswalks have its own deficiencies also. For example, there is no field in the target schema with an equal meaning, it may lead to unnecessary information fetched into the metadata schema. Also, only a handful of cases mapped equally in both metadata directions. This is due to the requirements and granularity of the data which each institution use. *Getty metadata crosswalk map* is an example of such single direction mapping (Woodley, 2008).

3.1.2. Resource Description Framework (RDF)

RDF (Resource Description Framework) is a "standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed" (W3C, 2014). However, RDF can be identified as a very formal and flexible technology capable of addressing a variety of problems which was developed under World Wide Web Consortium (W3C) specification. According to the RDF 1.1 Primer by W3C discusses reasons for using RDF as follows.

- Adding machine-readable information to Web pages enabling them to be displayed in an enhanced format on search engines or to be processed automatically by third-party applications.

- Enriching a dataset by linking it to third-party datasets.
- Interlinking API feeds, making sure that clients can easily discover how to access more information.
- Using the datasets currently published as Linked Data
- Building distributed social networks by interlinking RDF descriptions of people across multiple Web sites.
- Providing a standard compliant way for exchanging data between databases.
- Interlinking various datasets within an organisation (W3C, 2014)

The RDF data model is based on three core object types known as *Subject (Resource), Predicate (*Property) and Object (Literal). This is known as RDF triples (Figure 4). Through these triples can express any relationship and these triples can be connected.

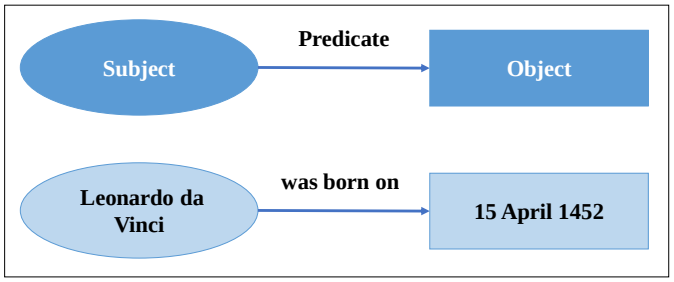


Figure 4: RDF triples example

RDF uses vocabularies such as FOAF, DC, SKOS, Schema.org and programming languages such as Turtle family of RDF languages, JSON-LD, RDFa and RDF/XML.

Many professional in various backgrounds use RDF technology to describe their resources and the CHI sector also utilises the same technology when dealing with web based semantical researches. The most prominent example is the Europeana Data Model (EDM) by the Europeana project. Europeana utilises the RDF graph and RDF Syntax (E.g. Turtle and RDF/XML) to describe their model. Figure 5 represents an aggregation of cultural heritage object with multiple digital representations on the web using a RDF diagram. According to the figure, EDM aggregates different web resources named as *edm:WebResource* through *ore:Aggregation*. Similarly, it has another resource called *edm:ProvidedCHO* which represents the real heritage object placed in a separate institutional collection. This type of relationship can be easily depicted through RDF graph.

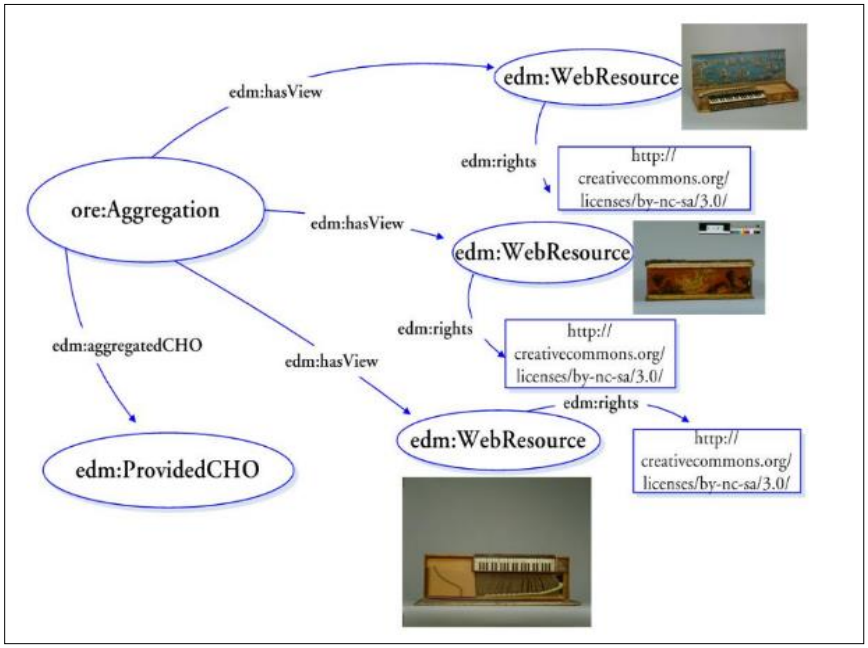


Figure 5: Europeana example showing cultural heritage object aggregation using RDF graph (Europeana, 2016)

3.1.3. Linked Open Data (LOD)

Linked Data refers to a set of best practices for publishing and connecting structured data on the Web. Technically Linked Data are data published on the web and they are machine readable, their external and internal links are well described etc. Technologies that support Linked Data are URIs (Uniform Resource Identifiers), HTTP (Hypertext Transfer Protocol), and RDF (Resource Description Framework) (Bizer et al., 2009). Berners-Lee (2006) who is a pioneer in web-based approaches describes a set of 'rules' for publishing data on the Web in a way that all published data becomes a part of a single global data space. These rules can be recognised as *Linked Data Principles*. "Linked Open Data (LOD) is Linked Data which is released under an open licence, which does not impede its reuse for free" (Berners-Lee, 2006). Anyhow, unlike Linked Open Data, Linked Data do not necessary to be open.

Since the focus is on online CHI, term LOD becomes an important factor when enriching CHI information related to this study. Nowadays museum-related institutions are also exploring and trying to publish their value information as LOD. For example, Getty vocabularies have developed as LOD while British museum provides its object information as Linked data.

Bore et al. (2012) presented an interactive methodology to ingesting and converting cultural heritage metadata as well as linking it to external data sources and publishing it as Linked Open Data. They used Amsterdam Museum metadata for this study and the system was supported by the ClioPatria semantic server.

Another research carried out by Knoblock et al. (2013). proposed an approach to developing a method for mapping museum data to a cultural heritage ontology and created tools for linking and validating the links to other sources. The aim of this study was to publish museum data as Linked Open Data related to the Smithsonian American Art Museum.

Kamura et al. (2011) proposed a system known as LODAC (Linked Open Data for Academia) Museum which enables to use Linked Data to integrate artistic and cultural fields that are naturally separated. LODAC Museum is an integrated metadata database of Japanese museum collections and they provide metadata in RDF formats. Basic functions of LODAC are scraping data from Web pages, mapping vocabularies, integrating unique items, publishing data as RDF etc.

3.2. Metadata Aggregation in CHI Environment

Simply metadata aggregation is, linking or connecting different metadata through their relationships. The previously discussed approaches such as crosswalks, RDF and LOD can be incorporated in metadata aggregation process. Swan & Awre (2006) in their research called *Linking UK Repositories*, outline the benefits of metadata aggregation as follows.

- Aggregations offer a breadth of access across many repositories, relieving end-users from accessing each one individually.
- Aggregations provide a single point of access to multiple sources of research and other materials to aid discovery.
- Aggregations offer an alternative route for enhancing metadata held within a repository.
- Aggregators can provide preservation and metadata enhancement capabilities to support the long-term storage of and access to the content etc.

Metadata aggregation is associated with many grounds such as library information, computer information, banking information, geographic information etc. This is a reliable approach to link information with diverse standards while share and enriching the same. Therefore, the same approach can be usefully assigned in CHI arena also. Related studies on Metadata aggregation in CHI environment can be discussed as follows.

Orgel et al. (2015) conducted a research and the title was *A metadata model and mapping approach for facilitating access to heterogeneous cultural heritage assets*. In this

paper, they focus on the definition of the metadata model and introduced a mapping approach and related tools to facilitate the use of heterogeneous cultural resources in EEXCESS (Enhancing Europe's eXchange in Cultural Educational and Scientific reSources) project funded by the European Union. The proposed metadata model is based on EDM and W3CPROV, and they introduced ontological mapping approach to map the information into the system.

Sugimoto et al. (2015) discussed a Manga Metadata Framework (MMF) in their article called Metadata in Cultural Context. They proposed a model based on FRBR and TV-Anytime to connect manga resources. There are two usage scenarios of MMF, such as Virtual bookshelf for manga and supporting reuse of products and bi-products during manga creation process. Since there were no controlled authority records for manga they used OCLC- Fiction Finder to aggregate bibliographical records. Then authors investigated few related databases such as Kyoto International Manga Museum (KIMM), Wikipedia and DBpedia to identify the FRBR Work instances for Manga. One limitation of this research is that the granularity of FRBR-Work is defined by the granularity of the bibliographic description instead of the content of the book. The authors further suggest that connecting metadata will enable new access points to contents. For instance, a manga which contains a historical story may be linked to cultural resources in Europeana data portal etc.

Zapounidou et al. (2014) examined how bibliographic data can be successfully aggregated with third party services such as cultural heritage portals. They tried to aggregate BIBFRAME source data using the Europeana aggregators through EDM library data alignment report. From BIBFRAME and EDM mapping they recognised that the expression of the BIBFRAME conceptualization in the Europeana framework using EDM classes and properties is achievable without significant loss of semantics.

Signore (2008) discussed on CHI on the semantic web. In this study, he argues that metadata level aggregation is not enough to fulfil the current trends and it should be replaced by core ontological approach. He further lists down issues related to CHI and related applications and then tries to introduce ontological and semantic web approaches for information integration.

DPLA (Digital Public Library of America) is another well-known example for metadata aggregation. DPLA aggregates existing metadata from libraries, archives, and museums to enable users to search and find collections and individual items. The resources of the DPLA are varied from print to digital media etc. they provide single point of access to millions of resources around the United States. In addition, DPLA-API provides access to metadata of the resources and all these data are freely available to the users (Guthro, 2013).

Europeana is a large data portal dedicated to aggregated, enrich and disseminate digital cultural heritage across memory institutions in the European Union. The official Europeana portal was launched in the year 2009 and currently, over 3,000 institutions across Europe contribute their resources to Europeana. Europeana portal is based on Europeana Data Model (EDM) which supports and manage the functionality of the system. The First data model of European was Europeana Semantic Elements (ESE) and now it has been further improved by the EDM. According to the Europeana Data Model Primer, "EDM is not built on any particular community standard but rather adopts an open, cross-domain Semantic Web-based framework that can accommodate the range and richness of particular community standards such as LIDO for museums, EAD for archives or METS for digital libraries" (Isaac, 2013, p. 6). EDM uses RDF graphs to describe its model and it utilises namespaces such as OWL, DC, SKOS, FOAF etc. The EDMs' core resources are the *Provided Cultural Heritage Object* termed as *edm:ProvidedCHO* and its related digital views known as *edm:WebResource*. All other relationships are based on these main core class resources and the metadata terms such as concepts or place names are semantically enriched by a range of Linked Open vocabularies (Isaac, 2013).

Though, this is a typically related study on cultural heritage resource aggregation it has major differences when compared with this present study. For example, the EDM model possesses a well-planned, top to bottom hierarchical approach. As identified earlier the data are packaged and provided by different providers. In addition, EDM model is mainly a resource based approach and the resources are confined to web resources only. However, the approach proposed by this research can be considered as a bottom-up approach because this tried to make relationships through metadata levels and it is not based on a pre-defined provided data sets. However, well-developed systems also have their own limitations and deficiencies. According to Peroni et al. (2012), complete integration of European resources is not fully done through EDM aggregation and this leads to losing of information. Similarly, they further describe the issues related to multilingual searching, semantic enrichment issues etc. Somehow by this time Europeana has achieved many of these issues and now they try to further improve EDM defining more avenues.

4. Research Methodology and Related Techniques

The methodology, data and related techniques used in this study can be discussed as follows.

4.1. Study Area of the Research

The study primarily focused on Sri Lanka which belongs to South Asian region. As identified in Chapter 2, despite its size Sri Lanka holds a rich cultural heritage unique to its community. Since this study discusses about the Sri Lankan CHI in abroad, the study area extends to Europe and the USA as well. Sri Lankan CHI in Europe is investigated through British Museum and Victoria and Albert Museum in London, UK while USA information is collected through Metropolitan Museum of Art in New York.

4.2. Materials, Data & Information used in the Study

The data and information used in the study can be mainly divided into online and offline information. Primarily, the information used in the literature and related bibliographical data are collected through online and printed mediums such as textbooks, journal articles and newspaper articles.

The most important CHI which was used to develop the current aggregation model was collected through above mentioned museum collections and they were extracted as online information. Since there were no Sri Lankan data online, had to rely on a printed museum catalogue to obtain the sample data. This printed catalogue was published by University of Peradeniya, Sri Lanka and it is related to a Museum attached to the Department of Archaeology of the same university (University of Peradeniya, 2004).

There are few reasons for specifically selecting British Museum and Victoria and Albert Museum in the UK for this study. First, most of the Sri Lankan artefacts are found in the UK compared to other museums in Europe. Moreover, since there is no language barrier in the UK, CHI information can be more reachable than other non-English European museums. As it was interesting to investigate CHI in different regions author selected Metropolitan Museum of Art in the USA as the third online collection. Metropolitan Museum of Art had a smaller Sri Lankan artefact collection compared to other two museums in the UK (Table 4).

4.3. Methodology of the Research

Figure 6 represents the general overview of the methodology used in this research. According to the figure first, data is extracted through four collections (BM= British Museum/ MM= Metropolitan Museum of Art / VA= Victoria and Albert Museum and SL= Sri Lankan Catalogue). The extracted data are used to create the mapping table and applicable metadata terms are identified according to the *Thematic*, *Spatial* and *Temporal* attributes. Through these identified and mapped CHI terms, the aggregation instances will be defined. The model is enriched through series of LOD resources and the final outcome will be a model to aggregated CHI of Sri Lanka.

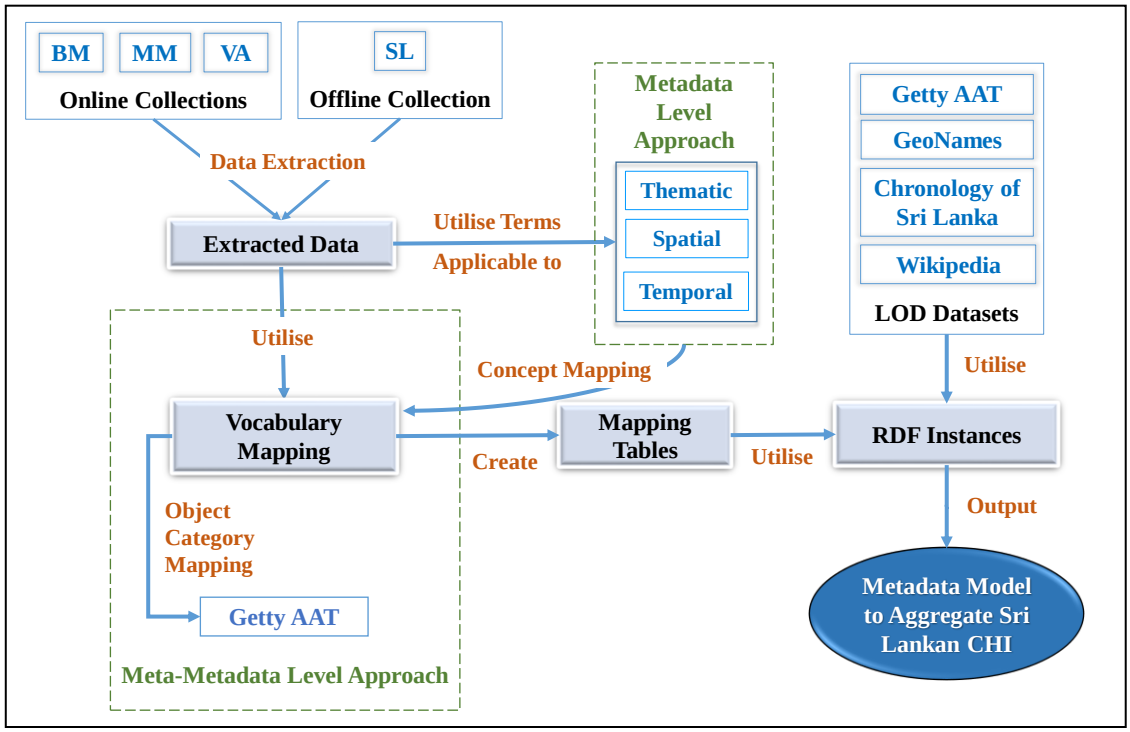


Figure 6: Methodology of the study

Next sections are dedicated to describe the above methodology in more detail manner.

4.3.1. Data Extraction

The advanced search options were used to perform queries in British Museum and Victoria and Albert Museum online collections. The advanced search options allowed to select the *production dates* and the criteria was given from 3rd century B.C. to 1975 A.D. aiming to avoid confusions with recent artefacts. Since there was a smaller amount of artefacts in

Metropolitan Museum of Art and there were no advance search options to perform criteria had to collect all related information of Sri Lankan CHI. In addition, to get local Sri Lankan data printed catalogue was used as follows (Table 4).

Table 4: Total number of records extracted from each collection

Museum Name	Abbreviated Museum Name	Total Records	URI/ References
British Museum, UK	BM	1779	http://www.britishmuseum.org/research/collection_online/search.aspx?place=41355&from=bc&fromDate=3&to=ad&toDate=1975
Victoria and Albert Museum, UK	VA	356	http://collections.vam.ac.uk/search/?listing_type=list&offset=0&limit=15&narrow=1&extrasearch=&q=&commit=Search&quality=0&objectnamesearch=&placesearch=sri+lanka&after=3&after-adbc=BC&before=1975&before-adbc=AD&namesearch=&materialsearch=&mnsearch=&locationsearch=
Metropolitan Museum of Art, USA	MM	123	http://www.metmuseum.org/search-results#!/search?q=sri%20lanka&page=1&searchFacet=Art
Senarat Paranavitana Teaching and Research Museum, Sri Lanka	SL	377 +	Prematilleke, P. L. (Ed.). (2004). The Catalogue of the Senarat Paranavitana Teaching and Research Museum. P. L. Prematilleke ed. Department of Archaeology, University of Peradeniya.

Searched online records were extracted by an open source application called Google Refine 2.5 (currently known as OpenRefine). Then the extracted records were transformed into Microsoft Excel format for further analysis and refinements. Sample records from the printed catalogue were typed into an Excel worksheet for easy manipulation (See Appendix I).

4.3.2. Metadata Mapping

Rather than working with just metadata level instances generating a meta-metadata level classes was an interesting approach in this study. The controlled terms used by each museum collections were used to fulfil that purpose. Here the main focus was only the *object categories* or *object types* defined by each museum and these terms were mapped into another top class vocabulary called Getty AAT (Art & Architecture Thesaurus). Getty AAT is a renowned structured vocabulary which consists of terms, concepts etc. related to art, architecture, archaeology, cultural heritage etc. This is a production by J. Paul Getty Trust and currently, it consists about 353,285 terms. Since Getty AAT is freely available as LOD it can be used as a noble resource for this kind of study (Getty, 2015). This formalisation was a key factor when aggregating terms related to spatial, temporal or thematic attributes of the CHI. The Mapping table related to the above mapping will be presented in chapter 5 (Table 6) (See Appendix II).

Other than the described mapping, metadata level mapping was done to make a formalisation between the spatial, temporal and thematic terms given by each museum. The study used DCMI (Dublin Core Metadata Initiative) terms as core metadata standards (Table 5). The main aggregating themes were *Spatial* (describes the production place/ origin of an object), *Temporal* (describes the production time/ period of an object) *and Thematic* (describes the concepts or themes of an object) terms extracted from the collections. *Identification* (museum identification no.) and *References* (bibliographical details related to the object) also took into the consideration as additional terms. The bracketed terms such as *(Title)* actually do not exist as attributes in the collections but the values given by the collection information derives the meaning of the attributes.

Since museums use different vocabularies to describe their collections, this kind of metadata level and meta-metadata level formalisation is essential for a clear understanding of the CHI in a diverse environment.

Table 5: Mapping selected museum terms to DC and DCMI terms

Aggregating Themes	DCMI Terms	British Museum	Metropolitan Museum	Victoria and Albert Museum	Sri Lankan Catalogue
Thematic	dcterms:title	(Title)	(Title)	(Title)	(Title)
	dcterms:subject	Subjects	Classification	Subjects depicted	Categorization of Objects
		Object type		Object	(Type)
	dcterms:description	Description	(Description)	Physical description	(Description)
				Historical context note	
Spatial	dcterms:spatial	Production place	Culture	Place of origin	(Place)
	dcterms:coverage	Findspot			
Temporal	dcterms:temporal	Date	Date	Date	(Date/ Period)
	dcterms:date		Period		
References	dcterms:relation	Bibliography	MetPublications	Bibliographic References	(References)
Identification	dc:identifier	Museum number	Accession Number	Museum number	No./ (Collection Code)

4.3.3. Model Development using RDF

With the aid of the formalised mapping tables, the relationships were developed over spatial, temporal and thematic terms of the cultural objects. At the same time, the concepts were enriched semantically through LOD resources such as;

a) Getty Art and Architecture Thesaurus (AAT)- for thematic terms
b) Wikipedia- for thematic terms
c) GeoNames- for spatial terms
d) Chronology of Sri Lanka (terms defined by the author) - for temporal terms

Aggregation instances and their relationships were depicted using RDF graph. Simple RDF triples were used to develop the model and it utilised the namespaces such as *dcterms, skos* and *rdf*. Figure 7 (*a*) and (b) illustrates a cultural object from Metropolitan Museum of Art. The related CHI of the same object can be described using RDF graph. According to the figure 7 (*a*) the cultural object is a *Buddhist statue* and its production date, production place, medium, dimension etc. can be identified using different terms unique to the Metropolitan Museum.

E.g. Object production date= *Period* and *Date*

Object production place= *Culture*

Somehow, using the mapping table (Table 5,) DCMI terms can be assign to Metropolitan terms and the result can be presented using a RDF graph (Figure 7 (b)). In RDF, oval shapes represent resources and rectangles represent values or literals. Similarly, the relationships and aggregations can be depicted using the same technique for multiple object information in different collections. This kind of links will be further discussed in chapter 5.

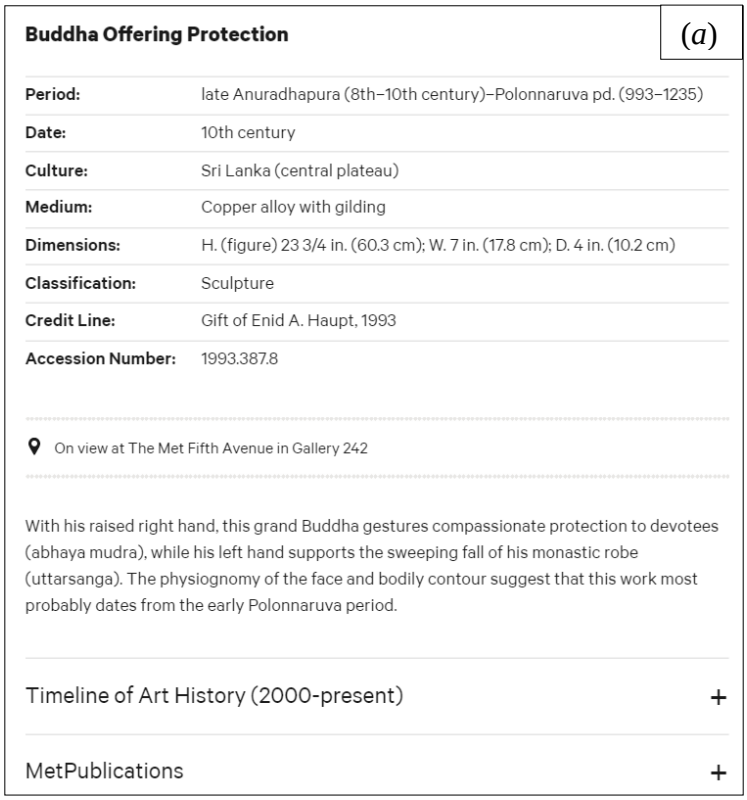


Figure 7 (a): A Museum object record in the Metropolitan Museum of Art collection

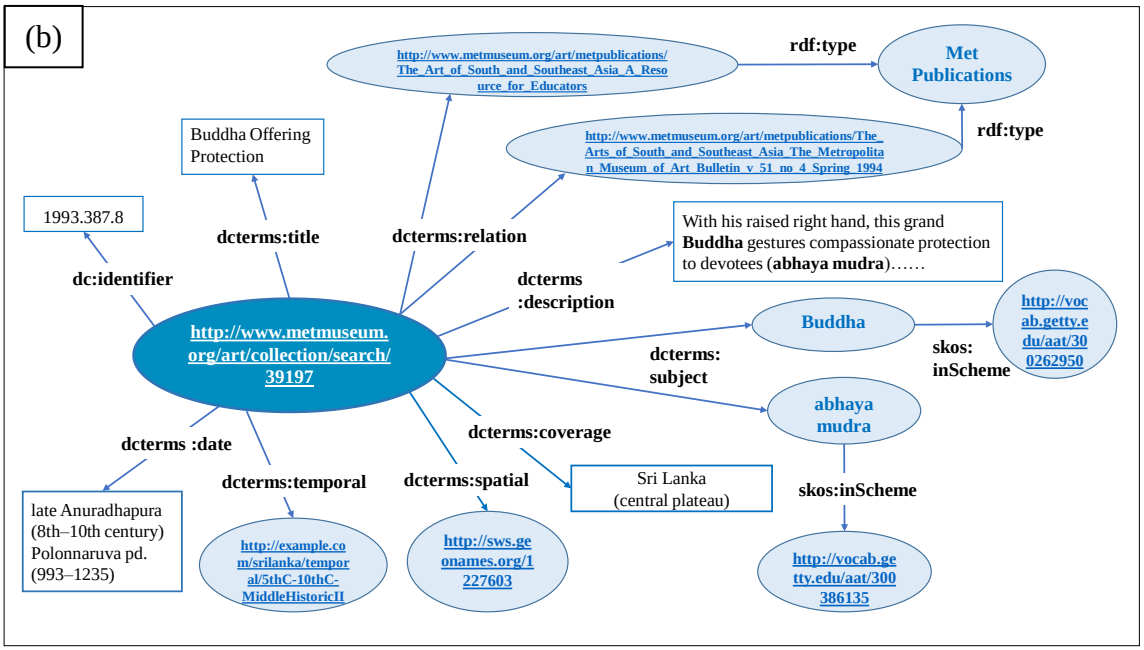


Figure 7 (b): RDF graph depicting the same information (selected attributes only)

5. Results and Discussion of the Study

5.1. Aggregation through Similarities and Dissimilarities

Aggregating metadata of diverse museum collections can be done through basic metadata level aggregation. Using simple DC mapping between specific museum terms (Table 5) an abstract formalisation can be established and through that mapping relationships can be created as follows.

Figure 8 & 9 represents few instances of museum CHI aggregation using a metadata model. Here the dark blue ovals represent the collection URLs while light blue ovals describe the concepts and their corresponding LOD resources. According to the Figure 8, similar cultural objects (*Buddhist sculpture*) from BM, MM, VA and SL were aggregated through equal concepts known as *Abhaya-mudra* and *Buddha* which are specific to Buddhist art. In parallel, temporal and spatial information also aggregated through similarities. For example, both SL and VA cultural objects have similar *object production dates* while all four museum objects possess a similar *production place* as well.

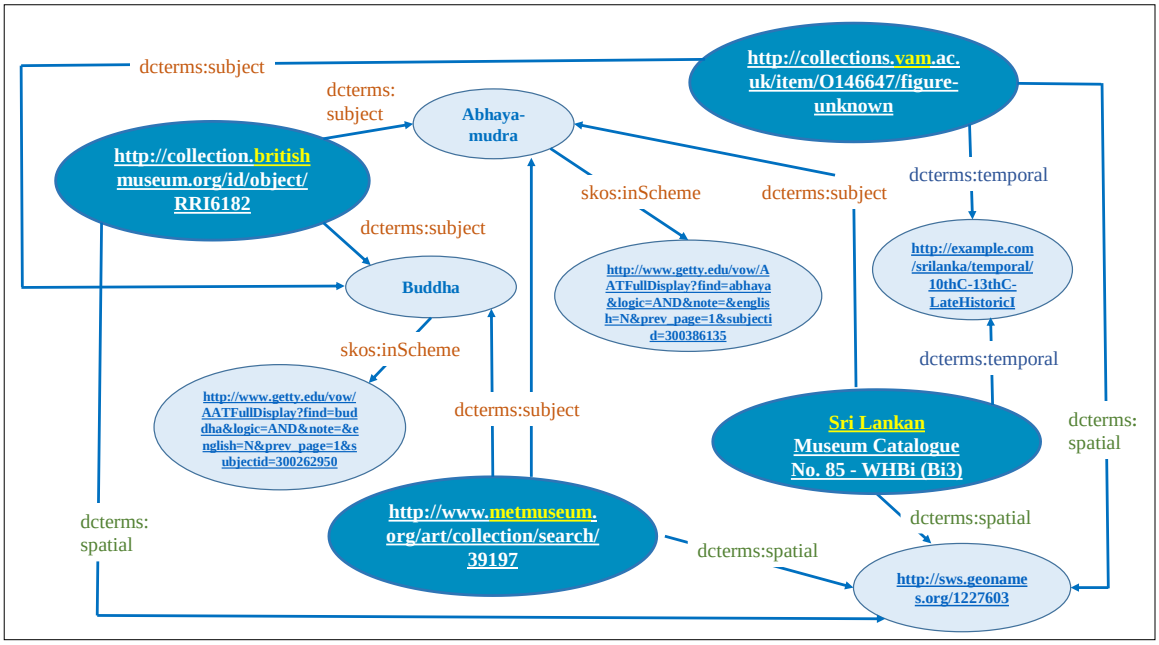


Figure 8: Aggregating similar CHI through metadata instances

Figure 9 shows an RDF example of dissimilar object aggregation using the same approach. Here aggregation object types are diverse. For example, BM has a *writing-slope*, MM has a *painted textile*, VA has a *figure* and SL has a *sculpture*. Somehow, all four objects represent the Hindu epic story *Ramayana*. Therefore, since they represent a similar concept the

metadata level relationship can be established as follows. The spatial and temporal relationships also can be established similarly.

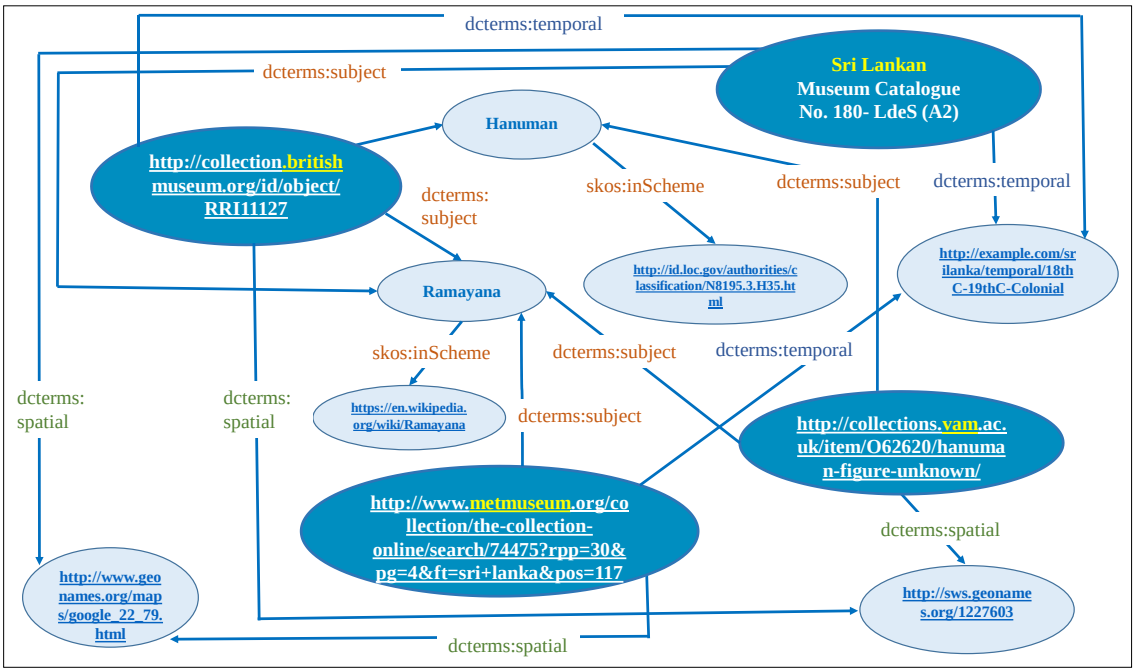


Figure 9: Aggregating dissimilar CHI through metadata instances

Somehow, above aggregation models are highly conceptual and it should be strengthened through a more solid level of crosswalk approach. Rather than aggregating through metadata level instances there should be a way to make a relationship between the CHI through more reliable formalisation. The next section of this chapter tries to seek a solution for that problem.

5.2. Object Category Mapping between Museum Vocabularies and Getty AAT

As introduced in the previous sections, formalisation between museums should be established prior to aggregation of metadata. In section 5.1, tried to introduce a very primary level aggregation model through metadata instances. However, real metadata aggregation model should have more solid and reliable approach and therefore, investigated more conceptual *object category* mapping to represent meta-metadata level CHI.

Here the *object categories* designated for Sri Lankan cultural objects by each museum were taken into consideration. Each museum had their own vocabulary to describe their object categories and in total, there were 285 object categories related to Sri Lankan cultural objects. From the identified terms, 240 terms were mapped into the targeted vocabulary (Table 6).

Table 6: Mapping between AAT and museum vocabularies

Colour key: Exactly match with AAT & similar across museums | Exactly match with AAT | Tentatively match with AAT

AAT		BM		MM		VA		SL	
Facet	**Concept**	**No.**	**Object Type**	**No.**	**Classification**	**No.**	**Object**	**No.**	**Objects Categories**
Objects	**containers (receptacles)**	1	bowl			1	Bowl		
		2	casket			2	Casket		
		3	Plate			3	Plate		
		4	reliquary			4	Reliquary	1	Reliquary
		5	amphora			5	Mug	2	Basket
		6	basin					3	Cosmetic Boxes
		7	box					4	Scabbards
		8	inkwell						
		9	jar						
		10	ewer						
		11	incense-burner						
		12	pill box						
		13	storage-jar						
		14	vessel						
		15	tobacco box						
		16	sheath						
		17	cosmetic vessel			6	Food vessels & Tableware	5	Betel Bags
		18	game bag					6	Chest
		19	gunpowder-flask					7	Chunam Containers
		20	lid box					8	Containers
		21	lime box					9	Medicine Boxes
		22	offering-bowl					10	Perfume Boxes
		23	pan					11	Powder Boxes
		24	rosewater-sprinkler					12	Trinket Boxes
		25	scent-fountain						
		26	textile bag (for manuscripts)						
Objects	**costume (mode of fashion)**	27	bracelet					13	Bracelets
		28	comb					14	Comb
		29	ear-ring					15	Ear-rings
		30	fan					16	Fan
		31	mask			7	Mask		
		32	pendant					17	Pendants
				1	Jewellery	8	Jewellery	18	Jewellery
		33	anklet					19	Armlets
		34	dance-mask					20	Bangles
		35	hair-pin					21	purse
		36	necklace						
		37	necklace bead						
		38	pendant necklace						
		39	ring						
		40	signet rings						
		41	dance-mask (kolam mask)						
		42	dance-mask sanni mask						

Objects	**architectural elements**	43	panel			9	Panel	22	Panel
		44	arch						
		45	door-jamb						
		46	Tile						
		47	plaque door-fitting			10	Architecture	23	Building Materials
		48	plaque panel door-fitting						
Objects	**exchange media (objects)**	49	coin					24	Coin
		50	banknote						
		51	proof coin						
		52	specimen banknote						
		53	token						
		54	coin lakshmi plaque						
		55	proof banknote						
		56	treasury note						
Objects	**sound devices (equipment)**	57	drum	2	Membranophone-double-headed / waisted drum				
		58	bell					25	Musical Instruments
		59	cymbal						
		60	gong						
		61	trumpet						
		62	gong-beater						
Objects	**components (objects parts)**	63	cigarette-card						
		64	Firearms Accessories						
		65	lid						
		66	stopper						
		67	book cover	3	Armor Parts			26	Book covers
		68	cover					27	Handle
		69	hip-wrapper					28	Knob
		70	hip-wrapper textile						
		71	perfume-bottle (stopper)						
		72	sculpture handle						
Materials	**materials (matter)**	73	ivories			11	Ivory		
		74	textile			12	Textile		
		75	tile			13	Tile		
		76	bones			14	Ceramics	29	Painted Cloths
		77	cloth						
Objects	**equipment**	78	comb					30	comb
		79	knife	4	Knife			31	knife
		80	ladle					32	ladle
		81	stylus					33	stylus
		82	hoe					34	Hatchets (Axes)
		83	plough					35	Spectacles
		84	yoke (harness)					36	staff
		85	fire-cover			15	Smoking Accessories	37	Areca Cutters
		86	plaque comb			16	Tableware & Cutlery	38	Betel Pounders
						17	Tools and Equipment's	39	Scrapers
Objects	**furnishing (artifacts)**	87	lamp					40	Lamp
		88	cressets			18	Furniture		
		89	throne						
		90	vessel-stand						
		91	tripod						
		92	writing-slope						

Objects	**weapons**	93	dagger	5	Daggers				
		94	sword	6	Sword			41	Sword
		95	knife dagger	7	Firearms				
		96	sheath knife						
		97	spear						
		98	sword sheath						
		99	knuckle-duster	8	Archery Equipment-Bows				
		100	valari	9	Shafted Weapons				
Objects	**recreational artifacts (equipment)**	101	dice					42	Dice
		102	game-board						
		103	drawing album						
		104	mancala-board						
Objects	**information forms (objects)**	105	document	10	Book	19	Posters	43	Banner
		106	flag						
		107	manuscripts						
		108	scroll						
		109	seal						
		110	photographic print (black and white) album					44	Copper Plate Inscriptions
		111	photographic print album						
		112	postcard photographic print (black and white) album						
		113	print album						
		114	print book- illustration						
		115	sealing						
		116	sketch-book painting drawing						
		117	stamp						
Objects	**visual works (works)**	118	drawing	11	Drawing	20	Drawing		
		119	figure			21	Figure	45	Figure
		120	oil painting			22	Oil painting		
		121	painting	12	Painting	23	Painting		
		122	print			24	Print		
		123	sculpture	13	Sculpture	25	Sculpture	46	Sculpture
				14	Photograph	26	Photograph		
				15	Metalwork	27	Metalwork		
		124	ivories			28	Embroidery	47	Geological specimens
		125	photographic print			29	Statue		
						30	Statuette		
						31	Watercolour		
						32	Woodwork		
		126	painting imitation	16	Textiles-Painted	33	Portrait miniature		
		127	shrine	17	Textiles-Painted and Dyed	34	Relief		
		128	slide 35mm (colour)						
Objects	**object genres (object classifications)**	129	weight					48	Weight
		130	ornament			35	Accessories		
		131	plaque			36	Ceramics		
		132	talisman			37	Personal Accessories		
		133	tool/ implement						
		134	animal remains						
		135	vegetal remains						
		136	vegetal remains painting						
Objects	**single built works (built environment)**	137	Shrine						
		138	Stupa						

As discussed in chapter 4, the Getty AAT vocabulary was used as the target vocabulary to establish this mapping and the resulted mapping table can be shown as above (Table 6). The definitions of the Getty terms used in the study can be found in the Appendix II. The mapped terms were categorised according to their level of similarities. Through that, some relationships were identified and the corresponding result can be described using RDF graphs and Turtle syntax as follows. The Figure 10 (*a*), (b), (c) and (d) describe few instances which represent the relationships with the AAT. Primarily, *bm, mm, va, sl* refers to the four museum collections and here the consideration was on the vocabulary terms only. The namespace *aat* refers to the Getty AAT which is the main resource of this mapping. These namespaces are described as prefixes under the Turtle syntax. The dark blue coloured ovals represent vocabularies and the light blue once represents the classes derived from the vocabularies. The relationships were described through *skos* terms as below.

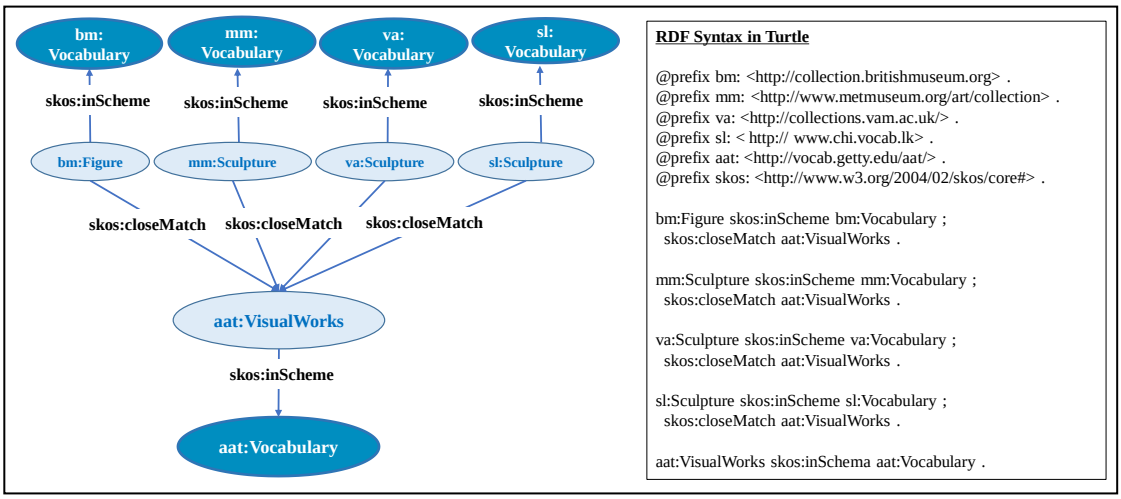


Figure 10 (a): RDF graph and syntax showing the close match relationship

Figure 10 (*a*) shows an instance of close match relationship and according to that *mm, va, sl* vocabularies use the term *Sculpture* while *bm* use the term *Figure* frequently. Although the meanings of the terms defined by individual museums are diverse technically, all the museums are addressing a similar object category. Therefore, terms *Sculpture* and *Figure* can be mapped into the broad term *Visual Works* in the *aat*.

Figure 10 (b) represents an instance of close and exact match relationship. Here the *bm* term *Figure* and *sl* term *Sculpture* can be exactly matched to the *aat* term *Sculpture* while *mm* and *va* term *sculpture* are closely matching with the same. This relationship is obvious when investigating the cultural object samples and definitions of the terms with the *aat* definition given to the term *Sculpture* under the *Object Facet*.

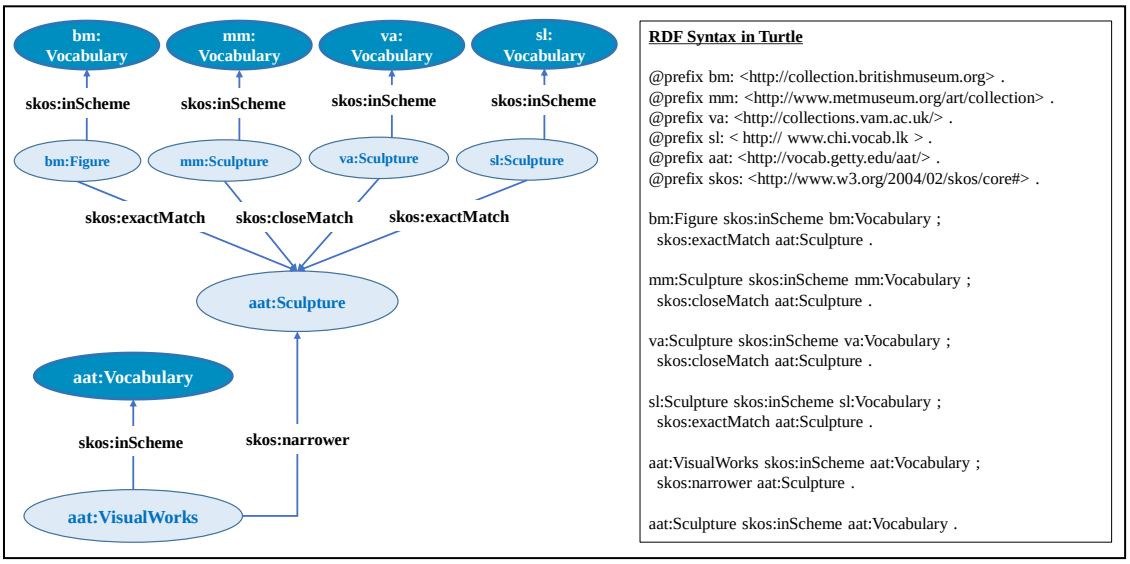


Figure 10 (b): RDF graph and syntax showing the close and exact match relationship

Figure 10 (c) shows another relationship. This time, the museum terms actually do not exist in the *aat*. However related synonyms can be found in the *aat*. Here *bm* term *Valari* which is a weapon similar to boomerang can be mapped to *aat*:*ThrowingSticks* category.

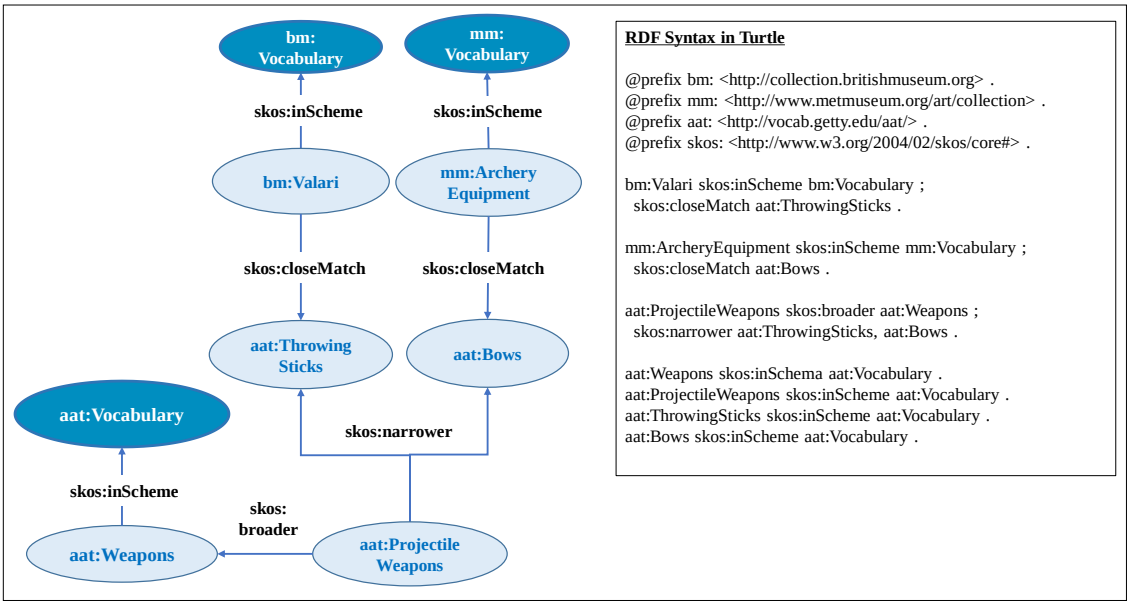


Figure 10 (c): RDF graph and syntax showing the close match relationship with different terms

In addition, *mm:ArcheryEquipment* can be a close match with *aat:Bows*. Somehow, all these terms can be categorised under *aat* term *Projectile Weapons* and finally into a broader term such as *Weapons*.

Figure 10 (d) shows another case derived from the above mapping. According to the figure, both *bm* and *sl* museums have the term *Comb* and the same term is visible in the aat also. However, *aat* categorises the term *Comb* under two broader terms known as *aat:HairOrnaments* and *aat:PersonalEquipment*. Therefore, according to Figure 10 (d), it is clear that some objects can have more than one meaning according to their usage.

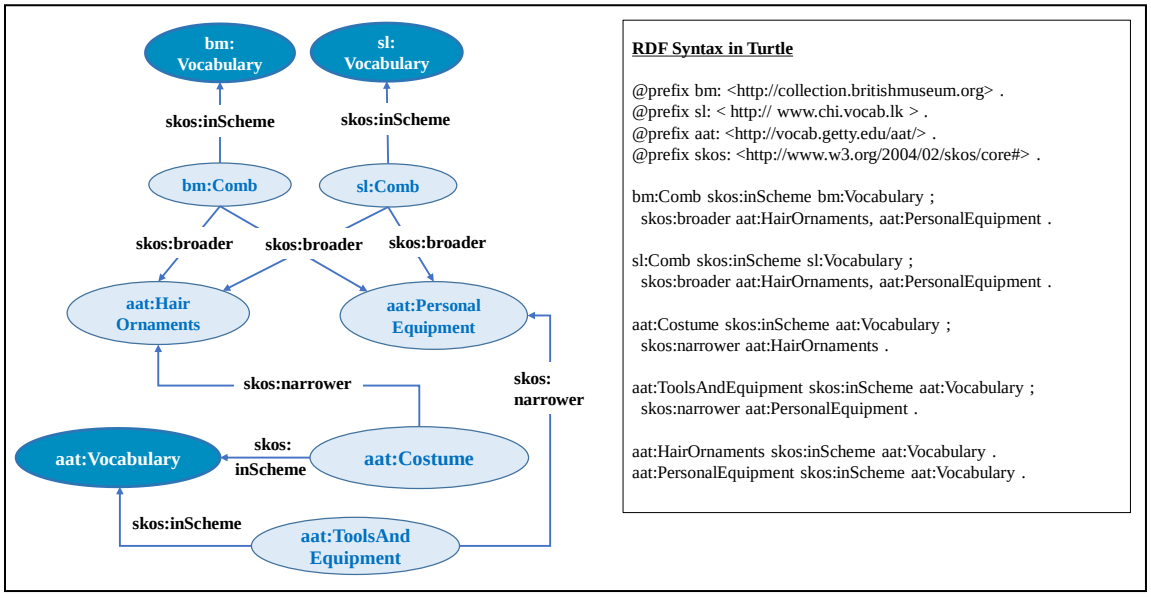


Figure 10 (d): RDF graph and syntax showing the broader and narrower relationship with different terms in different contexts

Similarly, objects can have diverse meanings according to materials used, concepts or themes represented, styles associated etc. Here in this study, the attention was only on the *cultural object types* or *categories* related relationships only.

5.3. Challenges Encountered and Limitations of the Research

The main challenge and the main research problem of this research were the diversity of the museum standards. This issue affects the aggregation approach in various directions. When it comes to mapping object categories into AAT, matching the meaning of the museum terms to AAT was the main challenge. The terms were defined according to museum requirements. As discussed above (Figure 10 (c)) some museums use different synonyms for the same object types. Therefore, had to investigate each and every term definition corresponded to each object

type to create the relationships. In addition, since four museum vocabularies had 285 object terms related to Sri Lankan cultural objects, out of that only 245 types were mapped to the AAT. As a result, had to omit 45 terms because those terms were explicitly defined for designated museums only. Since these terms created ambiguities assigning close or tentative terms was also not possible.

> E.g. BM object term such as *Mat Cooler* does not fit into specific AAT term. Instead, it can be mapped into *Mat* or *Cooler* separately which gives a different meaning to the designated object. In addition, SL terms such as *Makara-Torana* had no related term in AAT. *Makara-Torana* is a traditional Sinhalese decorative element and this type of local terms cannot be found in the AAT vocabulary.

Limitations of this research can be discussed as follows. Primarily, Sri Lanka had no online information and therefore, had to rely on a printed local museum catalogue. Yet this local museum catalogue represented a confined list of artefacts related to the medieval Sinhalese culture. Therefore, it was not a comprehensive list of object categories related to Sri Lankan cultural objects. Somehow, at this point the study had to depend on this offline data only. In addition, since there were no Sri Lankan made controlled vocabulary for cultural objects, identification of related terms was done using foreign museum vocabularies and other online resources such as Wikipedia.org. Finally, the study focused only on the English language based CHI only. There are renowned museums such as Rijksmuseum, Amsterdam, which possess a considerable amount of Sri Lankan cultural objects but their collection information was not readable because they are represented in the Dutch language. As a result, this study omitted the non-English museum collections and focused on English based collections only.

6. Conclusion and Future Direction

The study sought to find a suitable approach to aggregated CHI in the diverse environment and to fill the Sri Lankan CHI gap between memory institutions. To achieve that target, the study proposed a metadata level model as a base level CHI aggregation through spatial, temporal and thematic terms. Somehow, solid level metadata aggregation must have a concrete level formalisation between vocabularies. Aiming this target author suggested a crosswalk between four museum vocabularies with the Getty AAT vocabulary. The formalisation achieved through this *object category* mapping between museums can be used as the foundation to metadata aggregation. Also, the relationships derived from mapping was important to identify the cultural objects in different contexts with different granularities.

CHI aggregation is a popular and well-developed area and CHI users plus non-CHI users benefit from such approaches in numerous ways. However, CHI aggregation or CHI related metadata approaches are discussed very rarely in the Sri Lankan setting and are often an overlooked topic in the region. As discussed in chapter two, the DOCAM conference paper tried to investigate one such direction towards some specific CHI associated with cultural heritage monuments. This study narrowed down the broad theme *cultural heritages* to *museum CHI* and this time, it was towards the metadata aggregation. Somehow, investigating aforesaid fields which are not fully discovered yet would be a stepping stone to deliver CHI to the global arena in the Sri Lankan domain.

In future, the current research can be extended to develop a more comprehensive mapping between Sri Lankan cultural objects with renowned CHI vocabularies. Since lack of Sri Lankan based vocabulary related to cultural heritage was one of the main challenges that faced during this study, developing such vocabulary will be a crucial point to be addressed. Therefore, by obtaining more CHI specifically from Sri Lankan institutions and developing a complete list of authority terms solely for Sri Lankan cultural heritage will be the next step of this study. Since CHIs are not confined to a single institution MLA (Museums, Libraries, Archives) CHI aggregation is also another direction to be considered. Therefore, the same approach can be extended to aggregate more resources from diverse institutions in Sri Lanka as well as in South East Asia too. Since South East Asia shares similar cultural characteristics such as Buddhist or Hindu religion, writing script practices such as palm leaf manuscripts etc. developing an aggregation among those similarities to connect CHIs and make them available in LOD environment will be a good niche to be investigated in the future.

Acknowledgements

I would like to express my sincere gratitude to my principal supervisor Professor Shigeo Sugimoto for the useful comments, remarks, and the cooperation given through the learning process of this master thesis. Without his guidance and persistent help this thesis would not have been a reality. Furthermore, I would like to thank Professor Nagamori Mitsuharu and all the members of the Metadata Lab for their support given throughout my research.

A special thank goes to my colleague Honma Tsunagu for helping me to overcome technical difficulties and dedicating his time to teach me concepts related to this study. Also, I like to thank all the institutions for making their information available as open data to utilise for this kind of research without any difficulty.

I would like to thank my loved ones, who have supported me throughout the entire process, both by keeping me harmonious and helping me to achieve this target successfully. I will be grateful forever for your love.

Chiranthi J. Wijesundara

References

Baca, M. (Ed.). (2008). *Introduction to metadata*. Getty Publications. Online Edition, Version 3.0. J. Retrieved from http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.htm

Baca, M., Coburn, E., & Hubbard, S. (2008). Metadata and Museum Information. *Museum informatics: People, information, and technology in museums.* Marty, P. F. & Burton Jones, K. (Eds.). New York, NY: Routledge, 2, 107- 127.

Berners-Lee, T. (2006). Linked data. Design issues. *W3C*. Retrieved from https://www.w3.org/DesignIssues/LinkedData.html

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts.* USA: IGI Global, 205-227.

British Museum. (2016). *Collection online.* Retrieved from http://www.britishmuseum.org/research/collection_online/collection_object_details.aspx?objectId=247399&partId=1&searchText=sri+lanka&images=true&place=41355&from=bc&fromDate=300&to=ad&toDate=1975&page=1

Cabral, U. (2016). Activities of IFLA Preservation and Conservation (PAC) centre Sri Lanka, National Library and Documentation Services Board. *CDNLAO (The Conference of Directors of National Libraries in Asia and Oceania) Newsletter*, 85. Retrieved from http://www.ndl.go.jp/en/cdnlao/newsletter/085/855.html

Carrasco, L., & Vidotti, S. A. B. G. (2015). Dublin Core and CIDOC CRM harmonization. *International Conference on Dublin Core and Metadata Applications*, 98-200. Retrieved from http://dcpapers.dublincore.org/pubs/article/view/3779

De Boer, V., Wielemaker, J., Van Gent, J., Hildebrand, M., Isaac, A., Van Ossenbruggen, J., & Schreiber, G. (2012). Supporting linked data production for cultural heritage institutes: The Amsterdam museum case study. *In The Semantic Web: Research and Applications.* Springer Berlin Heidelberg, 733-747. Retrieved from http://oai.cwi.nl/oai/asset/19651/19651B.pdf

De Silva, K. M. (1959). *History of Ceylon- Volume II*. Sri Lanka: Ceylon University Press.

De Silva, P. H. D. H. (1975). *A catalogue of antiquities and other cultural objects from Sri Lanka (Ceylon) abroad.* National Museums of Sri Lanka.

Department of Archaeology. (2016). *Museums. Department of Archaeology - Sri Lanka.* Retrieved from http://www.archaeology.gov.lk/

Department of National Museums. (2016). Retrieved from www.museum.gov.lk

Eropeana. (2016). *MIMO and EDM*. Retrieved from http://pro.europeana.eu/page/mimo-edm

Farisz, H. & Dias, S. (2012, March17). National museum broken. *Daily Mirror*. Retrieved from http://www.dailymirror.lk/17497/-national-museum-broken

Gangabadadarachchi, V., & Amarasiri, M. S. U. (2009). Digital collection building initiatives of National Library and Documentation Centre. *Sri Lankan Journal of Librarianship and Information Management*, 2(1). doi: http://doi.org/10.4038/sllim.v2i1.440

Getty. (2015). *Art & Architecture Thesaurus Online*. Retrieved from http://www.getty.edu/research/tools/vocabularies/aat/index.html

Gill, T. (2004). Building semantic bridges between museums, libraries and archives: The CIDOC Conceptual Reference Model. *First Monday*, 9 (5). doi: http://dx.doi.org/10.5210/fm.v9i5.1145

Gilliland, A. J. (2008). Setting the stage. *Introduction to metadata.* M. Baca (Ed.). Online Edition, Version 3.0. Getty Publications. Retrieved from http://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.html

Guthro, C. (2013). Digital Public Library of America. Maine Policy Review. *Libraries and Information*, 22(1), 126-129. Retrieved from http://web.simmons.edu/~weigle/DIGCOLL/Guthro_DPLA.pdf

Harpring, P. (2010). *Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works.* Los Angeles, LA: Getty Research Institute. Retrieved from http://d2aohiyo3d3idm.cloudfront.net/publications/virtuallibrary/160606018X.pdf

Haslhofer, B., & Isaac, A. (2011). data. europeana. eu: The europeana linked open data pilot. *International Conference on Dublin Core and Metadata Applications*, 94-104. Retrieved from http://dcpapers.dublincore.org/pubs/article/view/3625

ICOM. (2007). ICOM statutes. *22nd General Assembly*. Vienna, Austria: International Council of Museums (ICOM). Retrieved from http://icom.museum/the-organisation/icom-statutes/3-definition-of-terms/#sommairecontent

ICOM. (2016). *ICOM missions.* http://icom.museum/the-organisation/icom-missions/

ICOMOS. (2002). *ICOMOS international cultural tourism charter: principles and guidelines for managing tourism at places of cultural and heritage significance.* ICOMOS International Scientific Committee on Cultural Tourism.

Isaac, A. (Ed.). (2013). *Europeana data model primer*. Retrieved from http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf

Kamura, T., Takeda, H., Ohmukai, I., Kato, F., Takahashi, T., & Ueda, H. (2011). Study support and integration of cultural information resources with linked data. *In Culture and Computing (Culture Computing)*, Second International Conference,177-178. doi:10.1109/Culture-Computing.2011.53

Knoblock, C., Szekely, P., Gupta, S., Manglik, A., Verborgh, R., Yang, F., & Van de Walle, R. (2013). Publishing data from the Smithsonian American art museum as linked open data. *In Proceedings of the 2013th International Conference on Posters & Demonstrations Track,*1035,129-132. Retrieved from http://ceur-ws.org/Vol-035/iswc2013_demo_33.pdf

Lor, P. J., & Britz, J. J. (2012). An ethical perspective on political-economic issues in the long-term preservation of digital heritage. *Journal of the American Society for Information Science and Technology,* 63(11), 2153-2164. doi: 10.1002/asi.22725

Museums Australia-Victoria. (2016). *Museum standards.* http://www.mavic.asn.au/museum_accreditation_program/museum-standards

NISO. (2004). *Understanding metadata*. USA: National Information Standards Organisation (NISO) Press. Retrieved from http://www.niso.org/publications/press/UnderstandingMetadata.pdf

Orgel, T., Höffernig, M., Bailer, W., & Russegger, S. (2015). A metadata model and mapping approach for facilitating access to heterogeneous cultural heritage assets. *International Journal on Digital Libraries*, 15(2-4), 189-207. doi: 10.1007/s00799-015-0138-2

Peroni, S., Tomasi, F., & Vitali, F. (2012). Reflecting on the Europeana data model. *Digital Libraries and Archives*. Springer Berlin Heidelberg, 228-240. doi: 10.1007/978-3-642-35834-0_23

Prematilleke, P. L. (Ed.). University of Peradeniya. (2004). *The Catalogue of the Senarat Paranavitana Teaching and Research Museum*. Department of Archaeology, University of Peradeniya.

Ranasinghe, P. (2015). Collection, preservation and provision access to palm leaf manuscripts in Sri Lanka. *Heritage as Prime Mover in History, Culture and Religion of South and Southeast Asia, Sixth International Conference of the South and Southeast Asian Association for the Study of Culture and Religion-SSEASR (Abstract).* Sri Lanka: Centre for Asian studies of the University of Kelaniya, 81. Retrieved from http://repository.kln.ac.lk/handle/123456789/8296

Signore, O. (2006). The Semantic Web and cultural heritage: ontologies and technologies help in accessing museum information. *Information Technology for the Virtual Museum*, Sønderborg, Denmark. Retrieved from http://www.weblab.isti.cnr.it/papers/public/itvm2006.pdf

Sugimoto S., Nagamori, M, Mihara T. & Honma, T. (2015). Metadata in cultural contexts – from manga to digital archives in linked open data environment. *Cultural heritage information: Access and management*. Ian Ruthven & G. G. Chowdhury (Ed.). UK: Facet Publishing.

Swan, A., & Awre, C. (2006). Linking UK repositories: Technical and organisational models to support user-oriented services across institutional and other digital repositories. JISC Study Report. Retrieved from http://www.webarchive.org.uk/wayback/archive/20140614201504/http://www.jisc.ac.uk/publications/reports/2006/linkingukrepositoriesreport.aspx

UNESCO. (1980). Intergovernmental committee for promoting the return of cultural property to its countries of origin or its restitution in case of illicit appropriation. *Statement presented by the Democratic Socialist Republic of Sri Lanka concerning the restitution of significant cultural objects from Sri Lanka.* Session 1. Paris. Retrieved from http://unesdoc.unesco.org/images/0003/000387/038748eb.pdf

UNESCO. (2016). Definition of the cultural heritage. *What is meant by "cultural heritage"?.* Retrieved from http://www.unesco.org/new/en/culture/themes/illicit-trafficking-of-cultural-property/unesco-database-of-national-cultural-heritage-laws/frequently-asked-questions/definition-of-the-cultural-heritage/

UNESCO. (2016). *Tangible Cultural Heritage*. Retrieved from http://www.unesco.org/new/en/cairo/culture/tangible-cultural-heritage/

Vellucci, S. L. (2011). Metadata and authority control. *Library resources & technical services,* 44(1), 33-43. doi: http://dx.doi.org/10.5860/lrts.44n1.33

W3C. (2014). *RDF 1.1 Primer*. Retrieved from. https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/

W3C. (2014). *RDF*. Retrieved from. https://www.w3.org/RDF/

Wijesundara, C., Sugimoto, S. & Narayan, B. (2015). Documenting spatial and temporal information for heritage preservation: a case study of Sri Lanka. *Proceedings from the Document Academy*, 2 (5). Retrieved from http://ideaexchange.uakron.edu/docam/vol2/iss1/5

Woodley, M. S. (2008). Crosswalks, metadata harvesting, federated searching, metasearching: Using metadata to connect users and information. *Introduction to metadata*. M. Baca (Ed.). Getty Publications. Online Edition, Version 3.0. Retrieved from http://www.getty.edu/research/publications/electronic_publications/intrometadata/path.html

Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2014). Integrating library and cultural heritage data models: the BIBFRAME-EDM case. *Proceedings of the 18th Panhellenic Conference on Informatics*,1-6. doi: 10.1145/2645791.2645805

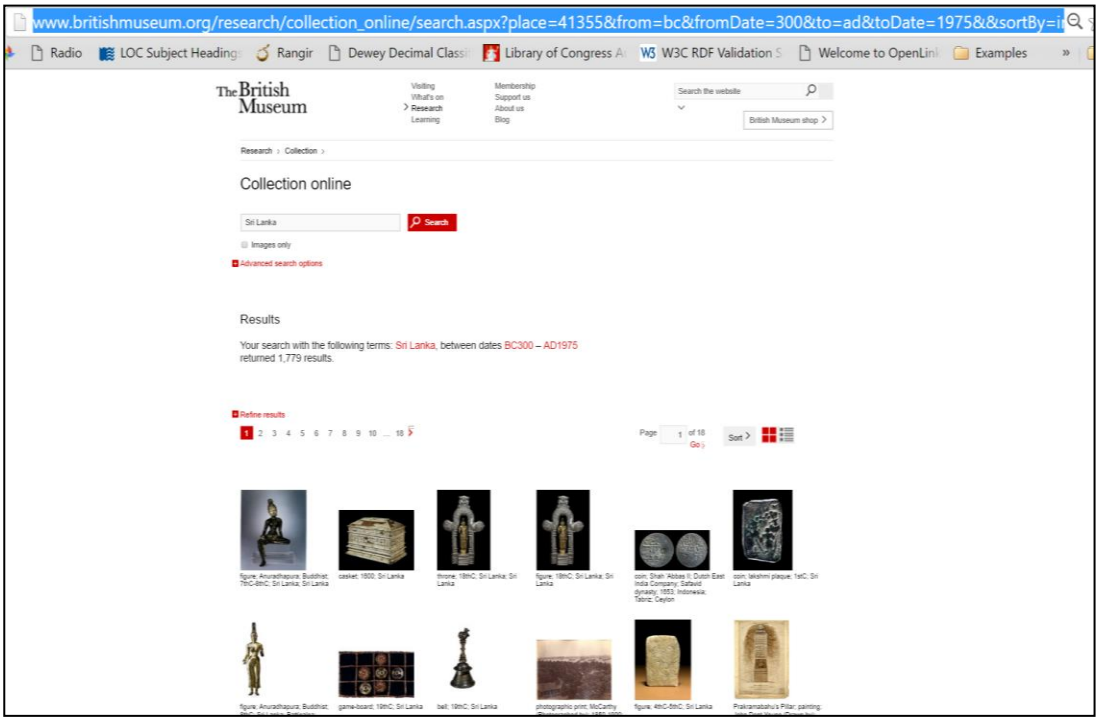
Zeng, M. L., & Qin, J. (2015). *Metadata* (2nd ed.). American Library Association.

Appendices

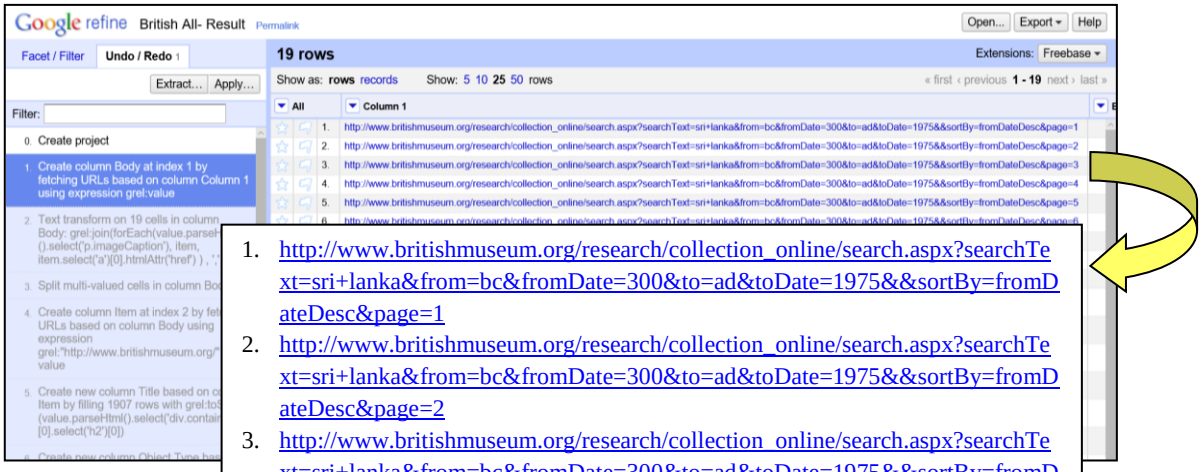
Appendix I: Process of extracting records from museum portals using Google Refine

1. Creating a new project in Google Refine

- **Search result of the British Museum Collection Online** for the Sri Lankan objects



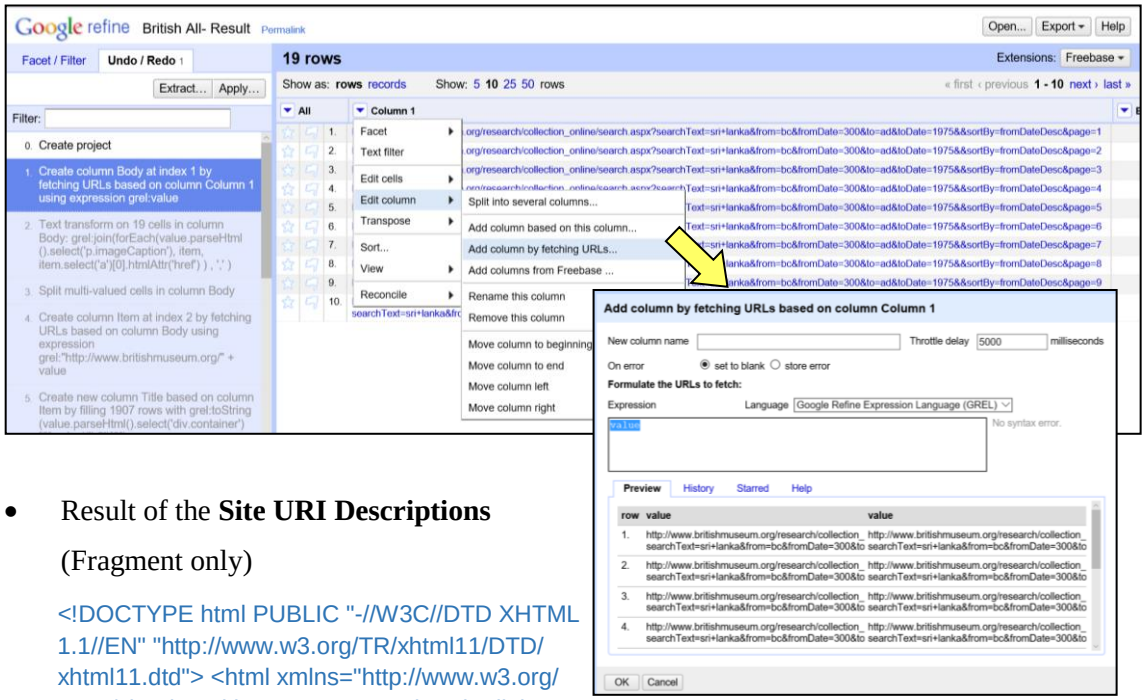
- **Search Result Page URIs** are used to create the **new project in Google Refine**



1. http://www.britishmuseum.org/research/collection_online/search.aspx?searchText=sri+lanka&from=bc&fromDate=300&to=ad&toDate=1975&&sortBy=fromDateDesc&page=1
2. http://www.britishmuseum.org/research/collection_online/search.aspx?searchText=sri+lanka&from=bc&fromDate=300&to=ad&toDate=1975&&sortBy=fromDateDesc&page=2
3. http://www.britishmuseum.org/research/collection_online/search.aspx?searchText=sri+lanka&from=bc&fromDate=300&to=ad&toDate=1975&&sortBy=fromDateDesc&page=3 …….

2. Extracting Site Level URI Descriptions

- **Column 1** (or **Search Result URIs**) are used to fetch **Site Level URI Descriptions**
- Use **Edit Column → Add column by fetching URLs** option on the **Column 1**



- Result of the **Site URI Descriptions**

  (Fragment only)

  <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN" "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd"> <html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en-GB"> <head><link href="/WebResource.axd?d=9SIonGbKeFVKuKsR_reirFktlYIIvlsRzc0rYI78h5uGJubjr5Gu1szq_pN7EXWmvegrJTmqH2vzyf_dhQ-I_UAyXIzm-Wxv3p4FHxzjSMgNMDm8qUyr1635tGEo3CFJxFWq-Dt6kspCudR5nNQ7MNEqVgI1&t=634383934428827289" type="text/css" rel="stylesheet" /><base href="http://www.britishmuseum.org/" /><title> British Museum - Collection search: You searched for sri lanka </title><meta http-equiv="Content-Type" content="text/html;charset=utf-8" />….

3. Extracting Item Level URIs from the Site URI Descriptions

- Use **Edit Cells → Transform** option on the **Site URL** to get the **Item Level URIs**.
- The expression is given in ***Google Refine Expression Language – GREL*** as follows.

  Join (forEach(value.parseHtml().select('p.imageCaption'), item, item.select('a')[0].htmlAttr('href')) , ',')

- Fragment of the output of the **Item Level URIs**

  /system_pages/beta_collection_introduction/beta_collection_object_details.aspx?objectId=252028&partId=1&searchText=sri+lanka&from=bc&fromDate=300&to=ad&toDate=1975&&sortBy=fromDateDesc&page=1,/system_pages/beta_collection_introduction/beta_collection_object_details.aspx?objectId=918902&partId=1&searchText=sri+lanka&from=bc&fromDate=300&to=ad&toDate=1975&&sortBy=fromDateDesc&page=1,/system_pages/beta_collection_introduction/beta_collection_object_de

tails.aspx?
objectId=3525911&partId=1&searchText=sri+lanka&from=bc&fromDate=300&to=ad&toDate=1975&&sortBy=fromDateDesc&page=1,/system_pages/beta_collection_introduction/beta_collection_object_details.aspx? ….

4. Splitting multivalued cells

- Splitting the above **Item Level URIs into Multivalued Cells**
- Use ***Edit Cells → Split Multivalued Cells*** option on the **Item Level URIs** to get the **output**
- Fragment of the output of the **Multivalued Cells**

/system_pages/beta_collection_introduction/beta_collection_object_details.aspx?objectId=252028&partId=1&searchText=sri+lanka&from=bc&fromDate=300&to=ad&toDate=1975&&sortBy=fromDateDesc&page=1
/system_pages/beta_collection_introduction/beta_collection_object_details.aspx?objectId=918902&partId=1&searchText=sri+lanka&from=bc&fromDate=300&to=ad&toDate=1975&&sortBy=fromDateDesc&page=1 ….

5. Fetching the HTML documents relevant to each item

- Use **Edit Column → Add column by fetching URLs** option on the **Site URI**
- The expression is

"http://www.britishmuseum.org/" + value

- Output of this will be another new column called **Item Column (which consist extracted HTML descriptions)**
- Fragment of the output of the **Item Column**

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN" "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd"> <html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en-GB"> <head><link href="/WebResource.axd?d=9SIonGbKeFVKuKsR_reirFktlYIIvlsRzc0rYI78h5uGJubjr5Gu1szq_pN7EXWmvegrJTmqH2vzyf_dhQ-I_UAyXIzm-Wxv3p4FHxzjSMgNMDm8qUyr1635tGEo3CFJxFWq-Dt6kspCudR5nNQ7MNEqVgI1&t=634383934428827289" type="text/css" rel="stylesheet" /><base href="http://www.britishmuseum.org/" /><title> British Museum - potter's tool/implement / dabber </title><meta http-equiv="Content-Type" content="text/html;charset=utf-8" /> ….

6. Extracting attributes from the Item Column (or Extracted HTML descriptions)

- Use **Edit Column → Add column based on column item** option on the **Item Column**

- The expression to get the **Title** of the item (**'h2'** refers to the **Title** of the item)

  toString(value.parseHtml().select('div.container')[0].select('h2')[0])

- The expression to get the **Description** of the item

  filter(value.parseHtml().select('ul.objectDetails')[0].select('li'), item, contains(toString(item), '<h3>Description'))[0].select('p')[0].htmlText()

- The expression to get the **Production Date** of the item

  filter(value.parseHtml().select('ul.objectDetails')[0].select('li'), item, contains(toString(item), '<h3>Date'))[0].select('ul')[0].htmlText()

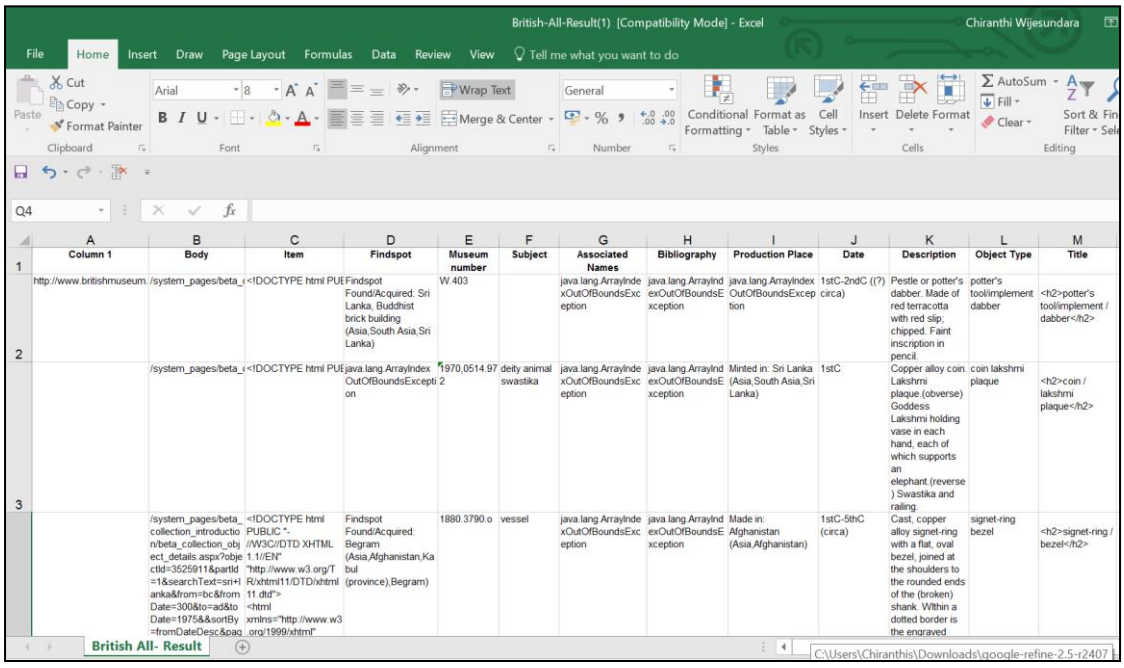
- The expression to get the **Production Place** of the item

  filter(value.parseHtml().select('ul.objectDetails')[0].select('li'), item, contains(toString(item), '<h3>Production place'))[0].select('ul')[0].htmlText()

- Other attributes such as **Reference No., Subject, Find Spot** etc. also can be obtained similarly

7. Exporting the final output

- The final result obtained from the Google Refine can be **Exported as an Excel file** and further **refinements** can be done through Microsoft Excel application
- Following Figure shows a screenshot of the exported Excel document



Appendix II: Definitions of the Getty vocabulary terms used in the study

No.	Facet	Concept	Definition (Note)	URI
01	Objects	containers (receptacles)	Receptacles or formed or flexible coverings designed to hold, store, or ship objects or substances	http://vocab.getty.edu/aat/300197197
02	Objects	costume (mode of fashion)	The mode or fashion of personal attire and dress, including the way of wearing the hair, style of clothing, jewelry, crowns, scepters, and other accessories of personal adornment, belonging to a particular nation, class, period, or special occasion, including all items worn or carried by people for warmth, protection, embellishment, or symbolic purposes. In English, generally expressed in the singular	http://vocab.getty.edu/aat/300178802
03	Objects	architectural elements	Forms, structural or decorative, developed originally or primarily as components of architecture, often adapted to other habitable spaces, such as in large vehicles, and often borrowed or imitated for structural or decorative use on other objects	http://vocab.getty.edu/aat/300000885
04	Objects	exchange media (objects)	Objects that are used in the exchange of goods and services and in the settlement of debts, and typically assigned a specific value	http://vocab.getty.edu/aat/300387350
05	Objects	sound devices (equipment)	Apparatuses, instruments, or other objects used to produce sound, whether musical or non-musical	http://vocab.getty.edu/aat/300387677
06	Objects	components (objects parts)	Constituent parts of a larger object. A component differs from an item in that the item can stand alone as an independent work but the component	http://vocab.getty.edu/aat/300241583

			typically cannot or does not stand alone. Examples are a panel of a polyptych or a discrete architectural component such as a dome	
07	Material	materials (matter)	The matter or substance from which a thing is or may be made; the tangible substance that goes into the makeup of a physical object. Physical substances, either naturally or synthetically derived, ranging from specific materials to types of material designated by their function	http://vocab.getty.edu/aat/300010358
08	Objects	equipment	Articles or physical resources used to array, dress up, rig out, equip, or otherwise provide with what is requisite for efficient action by a person or animal or for a thing	http://vocab.getty.edu/aat/300122241
09	Objects	furnishing (artifacts)	Works that facilitate human activity and to provide for physical needs of people in or around a building generally by offering comfort, convenience, or protection. An example of usage is in distinguishing the architecture of a building from the furnishings that are placed in and around it	http://vocab.getty.edu/aat/300037336
10	Objects	weapons	Implements or mechanisms used for defense or attack in combat, hunting, or war	http://vocab.getty.edu/aat/300036926
11	Objects	recreational artifacts (equipment)	Equipment and accessories used in any of a large array of activities that are engaged in for personal satisfaction or amusement during leisure time	http://vocab.getty.edu/aat/300218781
12	Objects	information forms (objects)	Types of textual, graphic, electronic, or physical items whose primary and original purpose is to record or convey	http://vocab.getty.edu/aat/300220751

			specific information. For forms in the sense of a document having blanks to be filled in, use "forms (documents)	
13	Objects	visual works (works)	Works of art and any objects that occupy space, are perceived by the sense of sight, and are created, rather than naturally occurring. Of special interest are those objects conveying a symbolic or expressive meaning or an aesthetic experience, although visual works are not limited only to such works. Visual works include pictorial and sculptural works, as well as time-based works such as performance art. They also include utilitarian objects of the type collected or valued by museums or individuals. Visual works do not include the performing arts	http://vocab.getty.edu/aat/300191086
14	Objects	object genres (object classifications)	Broad classifications for objects, which are material things that can be perceived by the senses; also includes electronic media.	http://vocab.getty.edu/aat/300185712
15	Objects	single built works (built environment)	Single structures in the built environment	http://vocab.getty.edu/aat/300004790